DOCUMENT RESUME

ED 190 219 PS 011 557

AUTHOR Chesterfield, Ray: And Others
TITLE An Evaluation of the Head Start Bilingual Bicultural Curriculum Development Project. Pilot Study Results and Child Assessment Measures.
INSTITUTION Juarez and Associates, Inc., Los Angeles, Calif.
SPONS AGENCY Administration for Children, Youth, and Families (DHEW), Washington, D.C.
PUB DATE 29 Jun 79
CONTRACT 105-77-1048
NOTE 150p.: For related documents, see PS 011 556-561, and PS 011 583.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Bilingual Education: Comprehension: Concept Formation: *Curriculum Evaluation: Influences: Language Acquisition: *Measures (Individuals): Multicultural Education: Parent Attitudes: *Preschool Curriculum: Preschool Education: *Program Effectiveness: Spanish Speaking: Tables (Data): Teacher Attitudes: Testing
IDENTIFIERS *Project Head Start

ABSTRACT

The purpose of this part of the multimethod Head Start bilingual/bicultural curriculum evaluation (Juarez and Associates, 1979), is to present the results of a pilot test of a battery of instruments to be used in assessing the effectiveness of four early childhood bilingual/bicultural curriculum models. The instruments pilot tested were the Bilingual Syntax Measure (second language acquisition), the CIRCO Subtests: Escuchen Este Cuento (Spanish language comprehension), the CIRCUS Subtests: Listen to the Story (English language comprehension), and the Preschool Inventory (English and Spanish concept development). Evaluation of the instruments focussed on three aspects: (1) the feasibility of successfully implementing the curriculum models in more than one setting: (2) the impact of each model on children enrolled in Head Start, on their parents, and on the Head Start teaching staff: and (3) the attitudes and opinions of Head Start staff and parents (who are in areas where the models are implemented) regarding the philosophy and objectives of each model and the quality of its programmatic contents. This document is concerned primarily with the second objective and presents a detailed analysis of the suitability of the field procedures and the appropriateness of the child impact assessment instruments. Also included are specific recommendations for improving both the procedures and the instruments during actual data collection. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED190219

Juárez and Associates, Inc.

An Evaluation of the Head Start Bilingual Bicultural Curriculum Development Project

Pilot Study Results and Child Assessment Measures

Authors: Ray Chesterfield, Helen Gonzales,
Thomas LaBelle, Luis Moll,
Ray S. Pérez, Marlene Zepeda

June 29, 1979

Contract No. HEW 105-77-1048

Project Director: Luis Moll, Ph.D.
Project Director: Ray S. Pérez, Ph.D.

Prepared for:

Dr. Steve Martinez, Project Officer
Administration for Children, Youth, and Families
Office of Human Development Services
Department of Health, Education, and Welfare

PS011557

1100 Glendon Avenue • Los Angeles, California 90024 • (213) 478-0826 • (213) 879-5899

Acknowledgments

A project of this scope, size, and complexity requires the creativity and hard work of many individuals. We would like to thank the following people for their contributions. Drs. Patricia Gándara and Richard Shavelson, who provided valuable technical assistance, and Mr. Jose Gonçalves, Ms. Helen Gonzales, and Ms. Myrna Valle, who helped in its preparation. Ms. Sandra Green and Ms. Dee Meyers, also contributed to the preparation of this report, and Ms. Margaret Mason gave editorial guidance.

CONTENTS

		Page
I.	Introduction	1
II.	Pilot Test Procedures	4
	A. Corpus Christi, Texas	4
	B. Rio Grande, Texas	4
	C. Milwaukee, Wisconsin	5
	D. New York, New York	5
	E. Sample Selection	6
	F. Selection of Local Examiners	9
	G. Procedures Used to Train Local Examiners	10
	H. Monitoring	14
	I. Test Administration	15
	1. Scheduling	15
	2. Facilities	16
	3. Order of Testing	16
	J. Debriefing	19
	K. Data Quality Control	20
	L. Problems Encountered	21
III.	Pilot Test Results	22
	A. Efficiency	22
	1. Language Check	22
	2. Language Comprehension -- Escuchen Este Cuento/ Listen to the Story	23
	3. Language Production -- Dímelo Tú/Tell Me You Say It	24
	4. Language Production -- Bilingual Syntax Measure	27
	5. Concept Development -- Preschool Inventory	27

B. Technical Adequacy of the Test Battery 29
1. Introduction 29
2. Language Dominance -- El CIRCO Language Check . 31
3. Language Comprehension -- Escuchen Este Cuento and Listen to the Story 38
4. Language Production -- Dímelo Tú and You Say It 50
5. Language Production -- Bilingual Syntax Measure 75
6. Preschool Inventory 83
C. Fairness 95
1. Environmental Conditions of Testing 95
2. Appropriateness 97
3. Difficulty 99
IV. Recommendations 101
A. Test Battery 101
1. Language Dominance 101
2. Language Comprehension 102
3. Language Production 104
4. Concept Development 111
V. Appendixes 120
A. Letter to Head Start Directors 121
B. Student Data Sheet 125
C. Student Description Characteristics 127
D. Job Description 132
E. Observational Checklist for Evaluation Team 134
VI. Bibliography 137

TABLES

		Page
1.	Characteristics of Sample Children by Site	7
2.	Summary Characteristics of Local Examiners	11
3.	Summary Data for the Language Check for English and Spanish Preferring Children	33
4.	Item Data for English and Spanish of the Language Check	34
5(a).	Summary Data for Escuchen Este Cuento for Spanish Preferring Children	39
5(b).	Summary Data for Listen to the Story for Spanish Preferring Children	40
6(a).	Summary Data for Escuchen Este Cuento for English Preferring Children	41
6(b).	Summary Data for Listen to the Story for English Preferring Children	42
7.	Item Data for Escuchen Este Cuento for Both English and Spanish Preferring Children	44
8.	Item Data for Listen to the Story for Both English and Spanish Preferring Children	45
9.	Internal Consistencies Reliabilities	48
10.	Reliability Estimates for Escuchen Este Cuento and Listen to the Story, by Language Group	49
11.	Summary Data for You Say It and Dímelo Tú	53
12.	Item Data for Dímelo Tú and You Say It.	55
13.	Reliability Estimates for You Say It and Dímelo Tú .	58
14.	Item Analysis Data for Dímelo Tú, Response Mode for Part IIa - Description	61
15.	Item Analysis Data for You Say It, Response Mode for Part IIa - Description	62
16.	Item Analysis Data for Dímelo Tú, Response Mode for Part IIb - Description	65

17. Item Analysis Data for You Say It, Response Mode for Part IIb - Description 66

18. Item Analysis Data for Dímelo Tú, Response Mode for Part IIIb - Narration 69

19. Item Analysis Data for You Say It, Response Mode for Part IIIb - Narration 71

20. Item Analysis Data for Dímelo Tú, Response Mode for Part IIIc - Follow-up Questions Narration 73

21. Item Analysis Data for You Say It, Response Mode for Part IIIc - Follow-up Questions Narration. 74

22. BSM Results by Language, Level, and Site 78

23. BSM - Syntax Acquisition Index (SAI) Item Analysis Mean Frequency Distribution Per Site 80

24. Summary Data for Spanish Preferring Children in Preschool Inventory 85

25. Summary Data for English Preferring Children in Preschool Inventory 87

26. Item Data: Spanish Preferring Children in Both Versions of Preschool Inventory 89

27. Item Data: English Preferring Children in Both Versions of Preschool Inventory 90

28. Estimates of Reliability for the Spanish and English Versions of the Preschool Inventory 92

29. Percent of Items Correct by Domain Component 100

FIGURES

Page

1. Test Measures and Testing Sequence for Pilot Study: Test Sequence A 17

2. Test Measures and Testing Sequence for Pilot Study: Test Sequence B 18

3. Correlation Matrix Comparing Each Instrument of the Test Battery by Language Group 36

4. Test Measures and Testing Sequence 114

I. Introduction

The purpose of this report (submitted in fulfillment of the requirements under Task 11, Phase II, of HEW Contract 105-77-1048, entitled "An Evaluation of the Head Start Bilingual Bicultural Curriculum Development Project"), is to present the results of a pilot test of a battery of instruments to be used in assessing the effectiveness of four early childhood bilingual bicultural curriculum models. The overall evaluation focuses on three aspects: (1) the feasibility of successfully implementing the curriculum models in more than one setting; (2) the impact of each model on children enrolled in Head Start, on their parents, and on the Head Start teaching staff; and (3) the attitudes and opinions of Head Start staff and parents where the models are being implemented regarding the philosophy and objectives of each model and the quality of its programmatic contents. This document is concerned primarily with the second objective and presents a detailed analysis of the suitability of the field procedures and the appropriateness of the child impact assessment instruments. Also included are specific recommendations for improving both the procedures and the instruments during actual data collection.

The following instruments were pilot tested:

- Bilingual Syntax Measure (second language acquisition)
- CIRCO Subtests: Escuchen Este Cuento (Spanish language comprehension)
  Dímelo Tú (Spanish oral production)
  You Say It (English oral production)
- CIRCUS Subtests: Listen to the Story (English language comprehension)
- Preschool Inventory (English and Spanish concept development)

The selection of the test battery used as measures of pupil growth was guided by the principle of finding the best fit between program objectives and tests. The procedure involved examination of tests in terms of technical adequacy, appropriateness for the target population, feasibility of administration, and congruence of the tests to curricular objectives (see Juárez and Associates, July 1978).

It is useful at this point to briefly review those characteristics which were identified as the most important attributes of tests to be used in the evaluation of preschool children in a bilingual (Spanish/English) setting. The pilot test was organized to provide empirical data on these two critical test features.

1. Efficiency - This refers to procedures for test administration, level of comprehension of the test, time of administration, cost, and availability. In short, instruments were examined to determine if they were appropriate and practical to administer. Included here are the time and administration constraints imposed by the field conditions. Data were collected on such issues as the length of the test battery and total administration time, sequence of testing, ease of administration by local paraprofessional testers, and ease of scoring the tests.

2. Technical adequacy - This refers to the validity and reliability of tests as reported by the test makers. Pilot test results are examined for reliability in a number of ways. Means and standard deviations are calculated for each test by site, language, and sex. Tests for differences were performed for each of these factors. An estimate of internal consistency was generated for each instrument. Items were compared to total test score to determine the utility of individual items. A correlation matrix was created to examine the relationship among tests. It is clear, however, that a test may be technically adequate, but that if the

objectives which it "measures" are not related to the curriculum objectives being assessed, and if the usability of the test in the field under severe time constraints is inadequate or is poor, the test (or tests) should be considered inappropriate.

In addition to the two main objectives of the pilot study, data were collected on what we have called the fairness of the test battery. Issues here include reactions to the testing environment by children and examiners, the appropriateness of the tests for the specific target population, and the difficulty of those test items which had been matched to cross model objectives (see Juárez and Associates, July 1978).

The body of this report is divided into several parts. Section II deals with the nature of the pilot study and the procedures used to carry it out. This includes descriptions of the four pilot test sites, the sample population, selection of the children, the examiners' characteristics, training procedures, and the test battery. Section III presents the pilot test results under three major headings: Efficiency, Technical Adequacy, and Fairness. Section IV concludes the report with a series of recommendations for alterations in individual instruments of the test battery and field procedures to improve pre-test data collection activities in the Fall of 1979.

II. Pilot Test Procedures

Four of the eight replication sites involved in the evaluation of the bilingual bicultural curriculum models were selected for participation in the pilot testing. In selecting these sites, one goal was to obtain sites that would provide a representative sample of the type of children who will participate in the bilingual curriculum programs starting in the Fall of 1979. Another goal was to obtain a diverse geographical representation of sites for the purposes of the pilot testing. During the selection of sites the curriculum model developers were consulted and given site selection criteria. They recommended one of their sites as most appropriate given the selection criteria. The four sites selected were located in Corpus Christi, Texas; Rio Grande City, Texas; Milwaukee, Wisconsin; and New York City, New York. A brief description of each site follows.

A. Corpus Christi, Texas

This site was situated in a semi-urban section of a moderate-sized city on the Texas Gulf coast. Two separate Head Start centers about one mile apart composed the site. One center had been in operation for three to four years whereas the other had opened fourteen months previously.

B. Rio Grande, Texas

This site was located in a small rural town in Southern Texas. The center was started twelve years previously in two permanent

bungalows on the grounds of a local elementary school, a locale it continued to occupy at the time of testing.

C. Milwaukee, Wisconsin

Located in a working class area on the west side of a large midwestern city, this site was situated in an abandoned church owned by the Archdiocese of Milwaukee. The center had been operating for seven years in a building which also served as the local community center.

D. New York, New York

This center, located on the Lower East Side of Manhattan, had been at the same location for nine years. The center was housed in the basement of a Dutch Reform church in the midst of a large housing cooperative populated by Blacks and Hispanics.

All four sites selected agreed to participate. Specific dates when the tests would be conducted were mutually agreed upon with each site's Head Start director. During initial phone contacts the goals of the pilot study were explained. Phone calls were followed by a letter explaining the pilot test procedures and giving dates; this letter was accompanied by student data sheets requesting background information on children, parental consent forms, and the examiner's job description (see Appendix A). Copies of these letters were also forwarded to the regional offices and to the curriculum model developers for their information and review.

We also requested the help of each Head Start director and his staff in recruiting at least twenty to thirty children per site who could participate in the pilot study, and in providing a list of potential examiners. From the student data sheets of these children a random sample was selected.

E. Sample Selection

The same procedures were followed for each site involved in the pilot study. In order to ensure that the psychometric properties of the tests and test procedures would be adequately assessed, the children were identified based on the following criteria: that children recruited be at least 4 years of age, that parental consent to testing be obtained, that children be from the classes in which the curriculum models were being implemented, and that the sample of children reflect the general composition of the Head Start center with regard to race/ethnicity and language preference. With only a few exceptions where expediency made it necessary to use children who did not meet the age criterion, these standards were adhered to consistently.

In addition to the twenty children who were selected randomly by Juárez and Associates for participation in the pilot study, five additional children were selected as alternates in case of absenteeism. The twenty children at each site were randomly assigned to the individual examiners.

A total of 97 children were selected for the pilot study. Table 1 summarizes the characteristics of the total sample. It should be noted that the average age of the children was 4 years and

TABLE 1

Characteristics of Sample Children by Site

Site	Mean Age	Sex		N
		Male	Female	
Corpus Christi	59.16*	8	16	24
English	59.23	6	15	
Spanish	58.66	2	1	
Rio Grande	58.56	13	12	25
English	57.00	1	0	
Spanish	58.62	12	12	
Milwaukee	54.55	7	11	18
English	55.54	2	9	
Spanish	53.00	5	2	
New York	57.80	11	10	21
English	57.94	9	9	
Spanish	57.00	2	1	

N = 88**

*In months

**Number of children with completed test folders.

7 months. Children ranged in age from 3 years and 6 months to 5 years and 7 months. Females outnumbered males 54 to 43.

Prior to testing, each child's language preference was assessed by three methods. First, the teacher's judgment of language preference was determined through the use of a short rating form measuring a child's language preference in various situations (see Appendix B). Second, the examiner's judgment of the child's use of language in different settings was elicited. Where inconsistencies existed, the examiner's judgment was accepted because of the contextual variety of assessment opportunities. Third, the EL CIRCO/ CIRCUS Language Check in Spanish and English* was used to verify the teacher and examiner judgments. If a child was unable to respond** in the designated language s/he was given the alternative version. These additional data led us to reclassify as either Spanish speakers or English speakers many of the children who had been categorized by the teachers as bilingual. As a result, 59% of the children were classified as preferring English, 38% as preferring Spanish, and only 3% as bilinguals for purposes of testing.

The ethnic and racial composition of the total sample of children tested was as follows: 68% of the children were Mexican Americans; 19% were Puerto Ricans; 6% were Black Americans; 3% were Anglo Americans; 3% were Dominicans; and 1% were Chinese. Therefore,

---

*The Spanish version was translated into English by Juárez and Associates.

**The criterion was 12 correct responses on the 16 item instrument with no more than 4 of those correct being in the second language.

the majority of the children tested in this pilot study were from Hispanic backgrounds (90%). The ethnic/racial composition and language preference of children differed by each test site. A summary of the characteristics of the children at each site is given in Appendix C.

F. Selection of Local Examiners

A job description for the examiner's position was developed and sent to the Head Start centers at each of the pilot sites (see Appendix D). The centers were given an opportunity to suggest appropriate people who would be interested in participating as examiners. Examiners who lived in the communities surrounding the four pilot sites were recruited to participate in the pilot study. Local examiners were selected because it was felt that these individuals would have a high degree of familiarity with the culture and language of the community in which the Head Start children lived. Because the oral language production measures required a written verbatim account of the child's responses, it was particularly important to select examiners who were bilingual and biliterate in Spanish and English. Consequently, a language screen was developed to assess each potential examiner's oral and written capability with the two languages. The screening procedures utilized by Juárez and Associates consisted of an interview to evaluate oral skills in Spanish and English and the administration of a test. This test required each potential examiner to write down a narrative word by word. The narrative was a tape recording made by a Juárez and Associates staff member of questions and answers from the

Bilingual Syntax Measure in Spanish and English. Each person's written response was then scored for grammatical correctness and accuracy.

A total of 19 bilingual and biliterate examiners were hired and trained to administer the entire test battery. Table 2 gives a composite of the educational level and previous work experience of the examiners at each site. It can be seen that approximately two-thirds of these examiners had a minimum of a high school diploma, with one individual having a Master's degree. Furthermore, at least 75% of the examiners had previous experience with young children, and some were parents of Head Start children. In fact, almost half of these examiners at one time or another had been teacher's aides in public schools or at Head Start centers. All of the examiners were female.

G. Procedures Used to Train Local Examiners

The examiner's training period ranged from one and a half to two and a half days. The length of training depended upon the experience of the local examiners, their demonstration of mastery of the testing procedures, and the practical constraints encountered at different sites. As part of the training, each examiner was given a manual which described each test and included samples of the test with instructions for administration, coding, and scoring procedures. Not all tests were, however, coded and scored by the examiners. The Bilingual Syntax Measure and Parts II and III of Dímelo Tú and You Say It were scored at Juárez and Associates offices in Los Angeles.

TABLE 2

Summary Characte.istics of Local Examiners

SITE	SEX		EDUCATION					WORK EXPERIENCE					
	M	F	No High School Diploma	High School Diploma	College or Univ. Courses	Bachelor's Degree	Higher Degree	No Experience	Teacher's Aides Sub. Teachers Volunteers	Experience Testing Children	Professional or Technical	Clerical	Health Services
CORPUS CHRISTI	0	4	1	1	0	2	0	1	3	1	1	1	1
RIO GRANDE	0	5	0	0	5	0	0	0	4	1	0	2	0
MILWAUKEE	0	4	2	0	2	0	0	0	3	0	0	1	0
NEW YORK	0	6	1	2	1	1	1	1	3	0	0	5	1
TOTAL	0	19	4	3	8	3	1	2	13	2	1	9	2

Additionally, pupil observation checklists and evaluation feedback checklists were provided, together with instructions on how to complete them. These checklists were designed to facilitate formative feedback from the local examiners to the Juárez and Associates staff concerning the adequacy of the training, children's reactions to the test situation, and the usability of the test battery in general. Information obtained from these checklists is reported in section III of this report.

The actual training sessions followed the language screening of the applicants. The goals of the training program were (1) to orient examiners to the role of testing young children, (2) to acquaint and familiarize examiners with the purpose of each test, (3) to train examiners in the procedures for administration of each specific test and the coding and scoring of certain tests, (4) to assess the recruitment and screening procedures for examiner selection, (5) to assess the utility of the examiner's training manual, and (6) to evaluate the overall training and testing procedures.

The training was divided into approximately four sessions ranging from two to three hours each. Although Juárez and Associates had developed a training plan (see Phase II: A Plan for the Pilot Study of Child and Parent Impact Measures, December 1978), each site required various adjustments owing to such unanticipated circumstances as delays in the arrival of testing materials, poor or inadequate training facilities, and examiner attrition.

The topics covered during each of the four sessions are outlined below.

1. First Session

- An introduction to the evaluation
- The role of the examiner
- Issues in testing young children
- Orientation to El CIRCO/CIRCUS Test measures for young Spanish speaking children
  - -- Language Check: administration and scoring procedures
  - -- Operating instructions for tape recorder

2. Second Session

- Introduction to oral language production measures
  - -- BSM: Spanish and English
  - -- Dímelo Tú and You Say It
- Administration and scoring of language comprehension measures
  - -- Listen to the Story
  - -- Escuchen Este Cuento

3. Third Session

- Concept Development: administration and scoring
  - -- Preschool Inventory: English version
  - -- Preschool Inventory: Spanish version
  - -- Assembling prop package
  - -- General review

4. Fourth Session

- Pupil Observation Checklist, including Tester's Log
- Evaluation Feedback Rating Forms

In addition to the sessions, examiners had an opportunity to interact with their assigned children and observe the children's language use in the classrooms. The purpose of this activity was to verify the teacher's judgments of the child's language preference, familiarize examiners with classroom environments, and facilitate rapport between examiners and their assigned children. As time allowed, examiners practiced administering portions of the test battery to each other.

H. Monitoring

A team of two staff members from Juárez and Associates conducted all prescreening and training of local examiners. These staff members were thoroughly familiar and experienced with the administration of all the tests in the battery. Observations were made of all local examiners during the first two days of testing (see Appendix E). The purpose of this monitoring was to ensure that all examiners were following procedures specified and explained during the training sessions, that all the necessary data were being collected, and to address any special problems that were encountered as the testing was initiated.

Before the Juárez and Associates staff members left the sites, a supervisor was selected from among local examiners. The supervisor's responsibilities were to oversee the remainder of the test administration, and to pack and ship test materials upon completion of the pilot testing to the Juárez and Associates offices in Los Angeles.

I. Test Administration

1. Scheduling

Once the examiners had been trained, they were introduced to the teachers in whose classroom the given curriculum model was being implemented. The four to six examiners per site were allowed some time to observe and interact with students in order to familiarize themselves with the children with whom they would be dealing and with the classroom procedures before testing actually began. This strategy allowed examiners to enter classrooms and remove children without disrupting ongoing activities.

Although it had been estimated that approximately one week would be needed at each site for administering the battery of tests to the sample children, such factors as student absenteeism, examiner attrition, and delays owing to special events (e.g., parties, field trips) impeded testing. As a result it took from eight to ten days to complete test administration at each site.

Examiners generally arrived at the Head Start facility between eight and nine o'clock and remained for five to six hours. A certain period of time was needed at the start of each day to arrange props, assemble toys, and lay out booklets required for the day's testing. Although total testing time for all of an examiner's children in any one day was two hours, the necessity of working within the daily schedule of activities (i.e., snack time, play time, and lunch) in which all students had to participate often required a four hour testing day. Upon completing testing assignments, examiners spent an additional one-half to one hour filling out rating forms on

the social and attending behaviors of the individual children with whom they had worked and on the overall testing environment.

2. Facilities

At each site one room was generally set aside for the pilot testing. As the room was an auditorium, an empty classroom, or an office, it was difficult to adequately separate the tester-student dyads. Although cloth screens were draped between examiners or children were placed back to back in an effort to reduce distractions with other children, there were some problems with interruptions by people entering the room, distractions with other children, and echoing or high noise levels which interfered with language taping at all sites.

3. Order of Testing

Tests were given in two different sequences during the pilot study. Sequence A (Figure 1), which was used with 45 children across the four sites, began with the Language Check and ended with the Preschool Inventory (PSI). Sequence B (Figure 2), given to 43 children, also started with the Language Check but reversed the administration of the other tests, with the PSI being given second and Dímelo Tú (You Say It) last. In both sequences, timing and pacing of the tests and test-wiseness were taken into consideration.

Each test session was limited to 30 minutes in order to minimize the effect of fatigue on a child's performance. This short span of time limited to two the number of tests which could be

Test Measures and Testing Sequence* for
Pilot Study
Test Sequence A

	Spanish Dominant		English Dominant	
	1st Language	2nd Language	1st Language	2nd Language
1st Session	Language Check (10 minutes)		Language Check (10 minutes)	
	Dímelo Tú (20 minutes)		You Say It (20 minutes)	
2nd Session		Listen to the Story (5-15 minutes		Escuchen Este Cuento (5-15 minutes)
	PSI Spanish (15 minutes)		PSI English (15 minutes)	
3rd Session		BSM-E (15 minutes)		BSM-S (15 minutes)
	Escuchen Este Cuento (15 minutes)		Listen to the Story (15 minutes)	
4th Session and Pick-up		PSI English (15 minutes)		PSI Spanish (15 minutes)
5th Session Pick-up Session				

*Testing sequence will be reversed for 50 percent of the children

Figure 2

Test Measures and Testing Sequence* for
Pilot Study
Test Sequence B

	Spanish Dominant		English Dominant	
	1st Language	2nd Language	1st Language	2nd Language
1st Session	Language Check (10 minutes)		Language Check (10 minutes)	
		PSI English (15 minutes)		PSI Spanish (15 minutes)
2nd Session	Escuchen Este Cuento (5-15 minutes)		Listen to the Story (5-15 minutes)	
		BSM-E (15 minutes)		BSM-S (15 mintues)
3rd Session	PSI Spanish (15 minutes)		PSI English (15 minutes	
		Listen to the Story (5-15 minutes)		Escuchen Este Cuento 5-15 minutes
4th Session and Pick-up	Dímelo Tú (20 minutes)		You Say It (20 minutes	
5th Session Pick-up Session				

*Session sequence will be reversed for 50 percent of the children.

administered during one sitting. Thus, a test session generally did not tax the attention span of a child.

As the pilot study tested all children in both Spanish and English and as several tests have very similar versions in both language , it was necessary to control for test-wiseness. Therefore, the sequences allowed for an interval of one to two days between the administration of the English and Spanish versions of the same instrument. In addition Sequence A was organized so that children were tested initially in their second language, since they would be capable of answering few questions and therefore would not be overly familiar with most items. Sequence A had the additional characteristic of presenting those tests which the pre-pilot identified as relatively easy first. No significant differences were found using analysis of variance between those children tested with Sequence A and those who experienced Sequence B. Evaluation checklists filled out by the test examiners showed that children's social and attending behaviors remained relatively high for all sites across the four test sessions, which suggests that the time periods were of appropriate length for children in the 4-5 year old range. In addition, in more than 90% of the cases for which data are available, examiners felt that the time span of the session was adequate or more than sufficient.

J. Debriefing

Prior to the departure of Juárez and Associates staff, a group debriefing session was held with the examiners. During the session, local examiners were given the opportunity to give information to the

Juárez and Associates staff about the adequacy of the training and their opinions regarding the difficulty of administering the test, as well as any other problems encountered during the testing. The data collected during the training and feedback sessions, in addition to the individual pilot test report completed by Juárez and Associates field staff, provided a basis for modifications in the training procedures implemented at the succeeding sites.

K. Data Quality Control

In addition to ongoing calibration of data collection procedures, checks aimed at ensuring the consistency of the data were built into the scoring and coding procedures. Coders were trained in one three-hour session by the same Juárez and Associates staff members who had trained examiners at the field sites. Once trained, coders examined each child's folder for completeness, then coded all responses left unscored by examiners, verified examiners' coding, and resolved questions. When the battery of a child had been completed, it was rescored and recoded by one of the trainers. If inconsistencies were found they were worked out jointly by the trainer and the coder. Trainers were then asked to give their impressions of the degree of agreement between themselves and the other coders for each instrument. These perceptions are discussed in part A of section III of this report, which deals with the efficiency of the measures in the test battery.

L. Problems Encountered

Several problems in addition to those mentioned earlier were encountered in the pilot study. Probably the most serious problem was that some of the training procedures were judged to be inappropriate given the experience and educational background of the local examiners selected. Specifically, wording of the test manual was too technical and complicated for use by the local examiners. Insufficient time was allocated for practicing and modeling each test, and the coding and scoring procedures were also judged as too difficult. A related problem was that of recruitment and screening. Our recruitment strategy did not provide us with a sufficient number of local people who were both bilingual and biliterate in Spanish and English. Other problems were that local examiners selected as supervisors were not familiar with shipping and packing demands. Therefore many of the materials were delayed in arriving at the Juárez and Associates offices on the specified dates, and several packages were misplaced en route to the sites via a commercial air carrier.

Specific procedures intended to improve the training and test administration to be used in the Fall 1979 pre-test are detailed in section IV of this report.

III. Pilot Test Results

A. Efficiency

Test efficiency refers to those factors or test attributes which make a test usable in the field. This section reports information collected for each test in the battery on the ease of test administration and the ease of scoring the test once the data are collected. For ease of administration the concern is with the test materials, their attractiveness to the preschoolers, and whether the children were responsive and attentive during the test. Other factors discussed are those related to the actual process of administration, such as the timing of each test and whether the test was culturally and linguistically appropriate for the preschoolers. Also included are practical problems the local examiners encountered in administering the test battery.

In addition, this section provides information regarding the ease of scoring the test. Factors that influenced the scoring procedures for each test, the training requirements for the scorers, the time needed to score each test, and the difficulty or ease in obtaining inter-scorer agreement are dealt with here.

1. Language Check

This brief instrument was reported by the Juárez and Associates staff and the local examiners to be easy to administer. The materials were colorful and attractive and included both pictures and props. The children were observed to be attentive and engaged with the task. During test administration it was important for the examiner to use words that were appropriate for the region in which the site is

located. It was also necessary to take care not to point to the correct picture or to indicate the correct response to the child in any other way. This test took an average of 6.7 mintues to administer.

The scoring and coding of the Language Check were relatively easy and inter-scorer agreement was good. It took approximately ten minutes to score and code, and verification of the scoring and coding revealed very few original errors.

2. Language Comprehension - Escuchen Este Cuento/Listen to the Story

In administering both Escuchen Este Cuento and Listen to the Story, some difficulties were encountered with unresponsiveness and inattentiveness on the part of the children. The loose story format and at times inappropriate vocabulary created difficulties for the children in following the sequence of story events and thereby in comprehending the story. As the illustrations were not of uniform attractiveness, children tended at times to pick attractive pictures prematurely, ignoring item content. The poor quality of graphics led to misinterpretations or competing, yet valid, responses by individual children. Materials lacked color and were therefore not engaging enough to help maintain attention already wandering because of lack of a consistent story line.

To ensure efficient administration of the test, it was necessary that the examiner be extremely careful to read the story slowly and clearly in order to hold the child's attention. Care also had to be taken by examiners to choose vocabulary that was appropriate to the locale and not to allow children to become so engrossed in marking pictures that they lost the story line.

Listen to the Story took an average of 10.9 minutes to administer to the Spanish preferring children and 14.4 minutes for the English preferring children. Escuchen Este Cuento, in turn, took an average of 9.3 minutes for the English preferring children and 13.4 minutes for the Spanish preferring children.

These two tests took approximately five minutes to score. They were both judged as easy to score, little time was needed to train the scorers, and inter-scorer agreement was considered good.

3. Language Production - Dímelo Tú/You Say It

As Dímelo Tú and You Say It are not direct translations of one another, but are tests with similar formats, item types, and scoring, the ease of test administration for each of the subtests is addressed separately.

The functional language section of Dímelo Tú was difficult for the children, although it held their attention. Graphics were not well designed or colorful and thus led to misinterpretation or valid alternative responses. Administration was difficult, as examiners were required to point to certain pictures at particular points in time while reading the test questions. Examiners had to be extremely careful to read items slowly and clearly and to use vocabulary appropriate to the region. Some confusion resulted, since responses had to be written verbatim and examples were given throughout the section. Thus, examiners at times could not keep their answer sheets in sequence with the questions being asked.

This subtest of Dímelo Tú took approximately 15 to 20 minutes to score. It was time consuming because the scorers had to rely on the examiners' notes, and subjective owing to the reliance on the

scorers' discretion. As a consequence of the problems of subjectivity, initial inter-scorer agreement was difficult. It was also observed that there were problems in training scorers because of the subjective nature of the test.

In the first description subtest (IIa) of Dímelo Tú, using a button, the children remained interested throughout. The cumulative nature of the test at times resulted in disadvantaging those children who failed to employ the appropriate vocabulary at the beginning of the test. As in other language production measures, difficulties were encountered by the examiners in capturing the children's responses verbatim. This subtest took approximately 5 minutes to score, and inter-scorer agreement was considered good.

The second description subtest (IIb), although short, tended to lose the children's interest as the objects used (shoelaces) were not especially engaging. In addition, the cumulative nature of the test presented the same difficulties as in subtest IIa above. Regional variations in the appropriate label for the shoelaces also created problems in the administration of the test. This subtest took approximately 5 minutes to score. Inter-scorer agreement was rated as good, although the possibility of distorting or misinterpreting the children's responses was present, since the scorers had to categorize responses in accord with established categories.

During the narration subtest, the children generally began to tire even though the graphics were colorful and engaging. Because examiners were unable to probe for responses, children had to initiate the discourse which, given their tiredness, slowed the administration. The procedures, which demanded that an examiner hold and point to a picture, read instructions, and write verbatim

responses simultaneously, also made the administration very difficult. The A portion of the narration subtest took approximately 5 minutes to score. Inter-scorer agreement was rated as good. For the B portion of the narration subtest approximately 10 to 15 minutes were needed for scoring. Inter-scorer agreement was lower as well as difficult to achieve, since the scoring required a high degree of knowledge of grammar.

With the follow-up to the narration subtest, confusion at times resulted because of the number of tasks which the examiners were asked to carry out simultaneously (i.e., writing, pointing, reading). Although the picture was again engaging, children were often inattentive at this point owing to fatigue. Approximately 5 minutes were needed for scoring. Inter-scorer agreement was rated as fair, since there were problems with certain items (numbers 3 and 5). There was also some difficulty in coding the responses of children who switched between languages.

The functional language subtest of You Say It generally failed to retain the children's attention because of its length, poor graphics which led to misinterpretation, and lack of color. Since the task closely paralleled the corresponding subtest of Dímelo Tú, problems were encountered in the time for scoring and with the answer sheet and its relation to the test questions and examples. Problems with other subtests were generally the same as those found with the descriptions and narratives of Dímelo Tú. The symbols and situations were able to retain children's interest depending on their nature, the graphics, and the timing of the test. The multiple tasks which the testers had to perform led to some degree of confusion and difficulty in administration, and care had to be taken to read

clearly and use appropriate vocabulary. The time needed to score these subtests is approximately the same as with the Spanish version, Dímelo Tú, and similar issues in scoring were also encountered. Both Dímelo Tú and You Say It took approximately 24 minutes to administer.

4. Language Production - Bilingual Syntax Measure

The Bilingual Syntax Measure was easily administered and its attractive illustrations, enjoyable cartoon-like characters, and informal, conversational nature tended to hold the attention of those children who knew the language that was used in the administration of the test. Although the test took only a short time to administer, the procedures had to be followed very carefully, as examiners were required both to write down all responses verbatim and to tape the conversations in order to increase reliability. This test took from 5 to 10 minutes to administer. Three systems of scoring were available. With all of the systems good inter-scorer agreements were possible to obtain, although a good understanding of the rules of grammar on the part of the scorer increased initial agreement. The most complex scoring system took approximately 20 minutes.

5. Concept Development - Preschool Inventory

Both the Spanish and the English versions of the Preschool Inventory proved to be easily administered. Not only was the instrument attractive enough to maintain a child's attention, but the toys and props provided items for manipulation and discourse. The variety of props, however, required that an examiner be very familiar with the objects used in testing and that the administration procedures be well organized. At times the use of common objects led to some

degree of test-wiseness on the part of children. It is important to investigate whether this would be particularly true for children who are somewhat bilingual. The test was time consuming to record, and it was difficult to score immediately the verbatim responses of the children. As with other tests, examiners had to read items slowly and clearly to avoid inattentiveness on the part of the children. The Spanish version of the PSI took an average of 9.9 minutes to administer to English preferring children and 20.3 minutes to Spanish preferring children. The English version took 7.7 minutes to administer to Spanish preferring children and 12.4 minutes to English preferring children.

Scoring of the Preschool Inventory took from 10 to 15 minutes. One difficulty was to decide how to score and code the responses of children who switch languages in answering a question. It was decided to score a response as incorrect or correct regardless of the language used. This was done in order not to penalize the children who could answer correctly but who did not use the language of the test. A notation was also made of the amount of language switching and with what items it occurred.

It was reported by the Juárez and Associates staff that it was relatively easy to train scorers for this test and that inter-scorer agreement was reached with little difficulty.

B. Technical Adequacy of the Test Battery

1. Introduction

In this part of the pilot study report the measures of the test battery are organized and described according to the domains each assesses. Questions addressed are (1) if the tests are appropriate for the representative sample selected; (2) if the psychometric properties of each instrument (item difficulty and discrimination, reliability, and relationship between tests) are acceptable; (3) if particular test results are similar to those obtained by the test developers working with like populations; and (4) if Spanish and English versions of the same test share certain characteristics (i.e., means, standard deviations, and correlations) which indicate that the versions are parallel. The statistical data presented here are used in conjunction with information provided under "efficiency" and "fairness" to determine when instruments and testing procedures are most appropriate for the evaluation of bilingual/bicultural Head Start curriculum models. These questions are also important for elaboration of a strategy for data analysis of the pre- and post-test (e.g., whether data should be separated by language group).

A number of tables are included to summarize the performance of the children who participated in the pilot study conducted during the months of February and March 1979. For each test a discussion of summary data is followed by an analysis of item difficulty and discrimination, the reliability of tests and

subtests, and the relationship between tests. Comparisons of pilot study results with test publishers' reported norms and comparisons between different language versions of the test given to the same group of children are made where appropriate. Information on the Spanish version of a test or subtest is generally given first.

In addition to the test specific investigations, analyses were performed to test for differences among sites, sexes, language preferences, and test order. As no differences were found, data were collapsed across those factors.

The summary data include mean scores and standard deviations for various subgroups. Only children who met the criteria on the Language Check (75% correct response rate) are included in the calculations. The criterion was imposed for two reasons: (1) to prevent the testing of children in a language in which they lacked sufficient skill and (2) to eliminate misleading low scores due to the child's lack of understanding of test instructions. The scores of children who did not complete the test battery for whatever reasons were also excluded from the calculations. Non-responses or refusals to respond to individual items of each test within the battery by children who completed the testing were coded as incorrect responses for the purposes of computing item difficulty, item discrimination, and reliability estimates.

The item data tables give the number of children choosing the correct response by language group, item difficulty, and the index of discrimination. The index of discrimination (point biserial correlation) is the extent to which an item differentiates

between children of high and low achievement. Achievement in this sense is defined as the child's score on the total test.

Reliability estimate tables for each test give the reliability coefficients for the total test, and when appropriate, for subtests for the pilot study sample and the norm sample. Estimates of the reliability give an idea of how accurately the tests are measuring the child's performance. The internal consistency coefficient alpha was used for this purpose. A correlation matrix was also generated to determine the relationship between measures in the test battery. The correlation coefficients provide information on the extent to which performance on one measure is related to performance on other measures of the test battery.

2. Language Dominance -- El CIRCO Language Check

The Language Check, which was administered prior to the other tests, was designed to identify children who did not have sufficient command of spoken language to understand the instructions for the various tests of the battery. The Spanish Language Check was translated into English by Juárez and Associates' staff, and both Spanish and English versions were used. The Language Check was administered to each child in his/her identified preferred language. Each version of the Language Check consists of 16 items that are individually administered. The items include questions and commands requiring both verbal and non-verbal responses.

a. Summary Data

Table 3 presents the means, total scores, and standard deviations for the CIRCO Language Check. Only scores of children who passed (achieved scores of 12 or greater with only 4 responses in the language other than that in which the test was administered) are included in the calculations. Consequently, the range of the test scores is restricted, as reflected in the standard deviations. Average scores were 14.24 (out of 16 items) for Spanish and 13.75 for English. Both versions of the Language Check were rather easy for the children in the pilot study, with children at the Rio Grande and Milwaukee sites scoring very close to the maximum. The Language Check proved an efficient screening instrument for children in the sample and also provided information as to an individual child's language skill.

b. Item Analysis Data

Table 4 presents response data for the Language Check. On the Spanish version, items 1, 2, 7, and 13 were answered correctly by all children taking the test, and the correct response rate on all items was relatively high. For the English version, at least 80% of the children answered items 1, 2, 5 through 12, 13, and 14 correctly. Even the most difficult item was correctly answered by 61% of the children. The items in both the English and Spanish version of the Language Check provided very little discrimination power owing to the restricted range of scores. Differences in item difficulty show no consistent pattern

TABLE 3

Summary Data for the Language Check

for English and Spanish Preferring Children

Sites	MEANS		S. D.		N	
	Spanish	English	Spanish	English	Spanish	English
Corpus Christi	13.25	13.35	1.50	1.80	4	20
Rio Grande	14.45	14.00	1.56	0.00	24	1
Milwaukee	14.57	15.09	1.27	1.45	7	11
New York	12.50	13.35	2.12	1.69	2	17
TOTAL for each Language Group	14.24	13.75	1.57	1.80	37	49

Number of items - 16

TABLE 4

Item Data for English and Spanish Versions of the Language Check

Spanish*	Proportion Answering Correctly	r_{bis}	N
LC 01	1.00	0.00	37.0
LC 02	1.00	0.00	37.0
LC 03	0.86	0.39	37.0
LC 04	0.45	0.08	37.0
LC 05	0.92	0.20	37.0
LC 06	0.83	0.15	37.0
LC 07	1.00	0.00	37.0
LC 08	0.97	0.03	37.0
LC 09	0.89	0.40	37.0
LC 10	0.97	-0.07	37.0
LC 11	0.97	-0.18	37.0
LC 12	0.84	0.34	37.0
LC 13	1.00	0.00	37.0
LC 14	0.84	0.23	37.0
LC 15	0.84	0.40	37.0
LC 16	0.81	0.30	37.0
English**			
LC 01	0.96	0.12	49.0
LC 02	0.96	0.25	49.0
LC 03	0.79	0.21	49.0
LC 04	0.65	0.31	49.0
LC 05	0.98	-0.01	49.0
LC 06	0.96	-0.01	49.0
LC 07	0.90	0.27	49.0
LC 08	0.96	-0.01	49.0
LC 09	0.90	0.13	49.0
LC 10	0.96	0.06	49.0
LC 11	0.96	-0.01	49.0
LC 12	0.61	0.16	49.0
LC 13	0.96	-0.07	49.0
LC 14	0.86	0.24	49.0
LC 15	0.80	0.31	49.0
LC 16	0.71	0.08	49.0

*For the Spanish version, alpha = 0.53.

**For the English version, alpha = 0.44.

across the two language versions of the test, nor were there any consistent patterns of item difficulty across the sites.

c. Reliability Estimates

Table 4 also presents the reliability estimate for each version of the Language Check with the alpha value for English equal to .44 and the value for Spanish equal to .53. The low correlations (.10 or less) between items 5, 6, 8, 10, and 11 and the total test score indicate that performance on these items of the English version is not highly related to the overall performance on that version of the Language Check. For the Spanish version, performance on items 4, 8, and 10 is not related to performance on the total test. The estimates of reliability were low as expected owing to the aforementioned restricted range of scores, and to the level of difficulty of the test items. Owing to the fact that the Language Check was the first test administered in both test orders, no order effects were expected.

d. Relationship of Language Check to Language Measures

The extent to which the Language Check served as a screen for the other language measures of the test battery was determined from a correlation coefficient matrix. Figure 3 presents correlation coefficients for each test. The upper diagonal of the correlation matrix presents the correlation coefficients of each of the measures for the English speakers across the sites. The lower diagonal contains the coefficients for the Spanish preference children. The margins present the mean scores and standard

FIGURE 5

Correlation Matrix Comparing Each Instrument of the Test Battery by Language Group

	ENGLISH											
LANGUAGE CHECK	YSI-I	YSI-IIa	YSI-IIb	YSI-IIIb	YSI-IIIc	YSI	PSIE	PSIS	LISTEN TO THE STORY	ESCUCHEN ESTE CUENTO	ENGLISH PREFERRING SCOR X	SD
LANGUAGE CHECK		.60	.20	.20	.26*	.51	.59	.66	.04	.51	.02	13.75	1.69
DT-I	.54		.28	.48	.41	.40	.89	.67	.10	.62*	.01	14.59	4.90
DT-IIa	.21	.26		.16	.20	.34	.45	.28	.16	.14	.06	3.53	1.24
DT-IIb	.19	.24	.35		.41	.27	.60	.34	.09	.28	.02	2.00	1.17
DT-IIIb	.16	.07	.17	.37		.46	.71	.44	.23	.37	.18	4.90	2.66
DT-IIIc	.48	.53	.45	.16	.03		.66	.49	-.01	.32	.01	5.29	1.82
DT	.56*	.83	.54	.51	.47	.70		.70*	.15*	.59*	.07*	30.31	9.10
PSIE	.28*	.32	.16	-.05	-.04	.28	.27*		.26	.57	.13	20.32	5.07
PSIS	.27*	.16	.43	.08	.01	.30	.26*	.13		.02	.83	12.88	7.41
LISTEN TO THE STORY	.22*	.19	.10	.10	-.03	.39	.24*	.37	.25		.11	13.00	3.50
ESCUCHEN ESTE CUENTO	.40*	.36	.10	.19	.03	.28	.34*	.31	.06	.23		9.31	5.42
	SPANISH											
SPANISH PREFERRED SCORES												
$\bar{X}$ =	14.24	12.84	2.91	1.62	5.59	4.57	27.54	11.13	20.35	7.62	13.46	
SD =	1.52	3.87	1.97	.95	2.26	2.00	6.78	8.46	5.81	5.37	3.93	
N =	37	37	37	37	37	37	37	33	37	37	37	

*P > .05

The upper diagonal presents English preferred children's performance and the lower diagonal presents Spanish children.

deviations for each test by language group. Only children completing the total test battery were included in this analysis.

For children who preferred to speak Spanish, performance on the Language Check was correlated moderately highly with performance on Dímelo Tú ($r = .56$) and Escuchen Este Cuento ($r = .40$). Figure 3 also presents data on English speakers, with performance on the Language Check related to performance on the other language measures such as You Say It ($r = .59$) and Listen to the Story ($r = .51$). The correlation between performance on the English Language Check and performance on Listen to the Story is slightly greater than the correlation between performance on the Spanish Language Check and on Escuchen Este Cuento. The Language Check for English speakers is also a slightly better predictor of the child's performance on the other measures of the battery than is the Language Check for Spanish speakers.

In summary, although the English and Spanish Language Checks have low discriminating power owing to the ease with which children correctly answered the items, they are effective instruments in determining whether children have sufficient language ability to understand the instructions and questions given in each of the instruments of the test battery. The language screen will be used in the pretest to aid in determining the language preference of the children for testing purposes.

3. Language Comprehension -- Escuchen Este Cuento and Listen to the Story

Escuchen Este Cuento and Listen to the Story were designed by the Educational Testing Service to measure the child's ability to comprehend a continuing discourse and to draw meaningful interpretations from this discourse. These two versions of the test contain scorable items connected to a story about children attending a circus.

The test is divided into two subtests. One measures simple listening comprehension skills and the other taps the ability to respond to instructions. Instructional response items are 1 through 6, 8 through 12, 14, 15, 22, and 23. Ten items -- 7, 13, 16 through 21, 24, and 25 -- are intended to produce responses based on the child's ability not only to comprehend the story and remember information but also to make inferences and interpretations based on what has been heard.

Each child who participated in the pilot study was given both language versions of the language comprehension measure. The children's responses were scored as correct, incorrect, or missing data.

a. Summary Data

Tables 5a and b and 6a and b summarize the data for both Spanish and English speakers on Listen to the Story and Escuchen Este Cuento. Spanish speakers' Escuchen Este Cuento mean scores range from a high of 17.3 at Milwaukee to 9.0 for Corpus Christi. For Listen to the Story the mean scores range from 10.25 at

TABLE 5a

Summary Data for Escuchen Este Cuento

for Spanish Preferring Children

Sites	Means	S.D.	N
Corpus Christi	9.00	2.94	4
Rio Grande	13.40	3.70	24
Milwaukee	17.30	4.80	7
New York	10.00	5.60	2
TOTAL	13.46	3.93	37
ETS' Norms Pre-kindergarten	13.37	3.93	363

Number of items = 25

TABLE 5b

Summary Data for Listen to the Story

for Spanish Preferring Children

Sites	Means	S.D.	N
Corpus Christi	10.25	1.71	4
Rio Grande	7.20	5.74	24
Milwaukee	7.30	5.46	7
New York	8.50	2.12	2
TOTAL	7.62	5.37	37

Number of items = 25

TABLE 6a

Summary Data for Escuchen Este Cuento

for English Preferring Children

Sites	Means	S.D.	N
Corpus Christi	11.55	3.14	20
Rio Grande	13.00	0.00	1
Milwaukee	10.60	9.01	11
New York	5.59	4.50	17
TOTAL	9.31	5.42	49

Number of items = 25

TABLE 6b

Summary Data for Listen to the Story

for English Preferring Children

Sites	Means	S.D.	N
Corpus Christi	12.20	2.68	20
Rio Grande	17.00	0.00	1
Milwaukee	17.90	3.11	11
New York	10.53	4.46	17
TOTAL	13.00	3.50	49
ETS' Norms	17.75	4.07	60

Number of items = 25

Corpus Christi to 7.2 at Rio Grande. From these mean scores it is evident that Escuchen Este Cuento and Listen to the Story were rather difficult tests for the pilot study children. The data also suggest that at Corpus Christi the preferred language of some children may have been identified incorrectly. Primary language mean scores represent slightly less than one half of the items answered correctly. For versions taken in the second language the mean score is one third of the items correct. The total mean scores of the children classified as English preferring were 13.0 for Listen to the Story and 9.3 for Escuchen Este Cuento. Once again the two tests were rather difficult for the children. The mean score represented slightly higher than one half of the correct items in the primary language and slightly less than one third of the correct items for the secondary language.

b. Item Analysis Data

Table 7 presents the percentage of Spanish and English preferring children who correctly answered any given item on Escuchen Este Cuento. Spanish preferring children's item difficulties ranged from over 91% correct for item 1, to 18% correct on item 18 on Escuchen Este Cuento. For English preferring children on Escuchen Este Cuento the item difficulties range from 57% of the children answering item 1 correctly to only 16% of the children being able to respond correctly to item 7.

As can be seen from Table 8, for Spanish preferring children on Listen to the Story the item difficulties ranged from 67%

TABLE 7

Item Data for Escuchen Este Cuento
for Both English and Spanish Preferring Children

	Proportion Answering Correctly	r_{bis}	N
English			
EC 01	0.92	-0.14	37.0
EC 02	0.59	0.15	37.0
EC 03	0.54	0.36	37.0
EC 04	0.54	0.45	37.0
EC 05	0.62	0.26	37.0
EC 06	0.78	0.38	37.0
EC 07	0.27	0.39	37.0
EC 08	0.70	0.24	37.0
EC 09	0.76	0.41	37.0
EC 10	0.40	0.48	37.0
EC 11	0.57	0.13	37.0
EC 12	0.38	0.50	37.0
EC 13	0.70	0.37	37.0
EC 14	0.62	0.26	37.0
EC 15	0.30	0.21	37.0
EC 16	0.43	0.06	37.0
EC 17	0.59	0.45	37.0
EC 18	0.19	0.44	37.0
EC 19	0.27	0.09	37.0
EC 20	0.49	0.30	37.0
EC 21	0.51	0.46	37.0
EC 22	0.35	0.02	37.0
EC 23	0.78	0.17	37.0
EC 24	0.67	0.29	37.0
EC 25	0.46	0.32	37.0
Spanish			
EC 01	0.57	0.52	49.0
EC 02	0.41	0.28	49.0
EC 03	0.39	0.40	49.0
EC 04	0.39	0.23	49.0
EC 05	0.41	0.29	49.0
EC 06	0.41	0.40	49.0
EC 07	0.16	0.36	49.0
EC 08	0.37	0.50	49.0
EC 09	0.53	0.69	49.0
EC 10	0.24	0.39	49.0
EC 11	0.47	0.66	49.0
EC 12	0.22	0.34	49.0
EC 13	0.51	0.57	49.0
EC 14	0.39	0.44	49.0
EC 15	0.18	0.37	49.0
EC 16	0.22	0.22	49.0
EC 17	0.55	0.62	49.0
EC 18	0.22	0.38	49.0
EC 19	0.24	0.38	49.0
EC 20	0.45	0.44	49.0
EC 21	0.49	0.56	49.0
EC 22	0.25	0.33	49.0
EC 23	0.47	0.51	49.0
EC 24	0.39	0.42	49.0
EC 25	0.37	0.50	49.0

TABLE 8

Item Data for Listen to the Story for Both English and Spanish Preferring Children

	Proportion Answering Correctly	r_{bis}	N
English			
LS 01	0.49	0.12	49.0
LS 02	0.67	0.40	49.0
LS 03	0.59	0.41	49.0
LS 04	0.51	0.34	49.0
LS 05	0.49	0.29	49.0
LS 06	0.75	0.32	49.0
LS 07	0.20	0.26	49.0
LS 08	0.59	0.22	49.0
LS 09	0.59	0.18	49.0
LS 10	0.37	0.32	49.0
LS 11	0.55	-0.01	49.0
LS 12	0.35	0.43	49.0
LS 13	0.69	0.22	49.0
LS 14	0.65	0.39	49.0
LS 15	0.45	0.16	49.0
LS 16	0.37	0.26	49.0
LS 17	0.75	0.27	49.0
LS 18	0.30	0.40	49.0
LS 19	0.51	0.06	49.0
LS 20	0.63	0.44	49.0
LS 21	0.49	0.34	49.0
LS 22	0.26	0.18	49.0
LS 23	0.63	0.14	49.0
LS 24	0.65	0.17	49.0
LS 25	0.53	0.50	49.0
Spanish			
LS 01	0.32	0.37	37.0
LS 02	0.67	0.18	37.0
LS 03	0.57	0.35	37.0
LS 04	0.19	0.11	37.0
LS 05	0.38	0.47	37.0
LS 06	0.54	0.64	37.0
LS 07	0.19	0.04	37.0
LS 08	0.35	0.41	37.0
LS 09	0.35	0.47	37.0
LS 10	0.30	0.39	37.0
LS 11	0.19	0.28	37.0
LS 12	0.19	0.40	37.0
LS 13	0.35	0.60	37.0
LS 14	0.24	0.39	37.0
LS 15	0.24	0.38	37.0
LS 16	0.24	0.38	37.0
LS 17	0.24	0.39	37.0
LS 18	0.24	0.31	37.0
LS 19	0.08	0.32	37.0
LS 20	0.22	0.41	37.0
LS 21	0.22	0.53	37.0
LS 22	0.16	0.30	37.0
LS 23	0.35	0.62	37.0
LS 24	0.46	0.69	37.0
LS 25	0.27	0.52	37.0

correct on item 2 to only 8% able to answer item 19. English speaking children's levels of difficulty for the same test range from 75% answering items 6 and 17 correctly to 20% responding correctly to item 7. Listen to the Story is the most difficult test of the battery, as determined by the range of item difficulty values. More than half the items were answered correctly by only 30% of the children. When the two subtests of Listen to the Story were examined separately, no consistent patterns of item difficulty were found for either of the subtests on Listen to the Story or Escuchen Este Cuento for the children categorized as English preferring. For the children categorized as Spanish preferring, interpretation and inference items on the average were much more difficult than the simple comprehension items. This finding was anticipated, since it is thought that interpretation and inference are skills not directly addressed in the particular programs in which the children found themselves.

Spanish preferring children did not do as well on the language comprehension test in their secondary language as in their primary language. In particular those items that required more than simple comprehension were very difficult for these children.

The index of discrimination calculated for both Listen to the Story and Escuchen Este Cuento was the point biserial correlation coefficient. The point biserial correlations were computed for each of the items in the test for English preferring and Spanish preferring children and are presented in Tables 7 and 8. As the test became more difficult for these children, the item discriminative

power increased. Since the values of the point biserial correlation fluctuated between .26 to .42, it can be said that these tests have acceptable to adequate discriminating power, as acceptable point biserial correlations for achievement tests generally fluctuate around $r_{bis} = .40$.

c. Reliability Estimates

Reliability of Listen to the Story and Escuchen Este Cuento was calculated using a measure of internal consistency (alpha coefficient, see Table 9). For each of the language measures the correlation coefficients were generated by language group. These estimates are presented in Table 10. Taking the total sample of children, the reliability of both Escuchen Este Cuento and Listen to the Story is moderately high, with r = .86 and r = .84 respectively. For each language group the estimates of reliability are somewhat modest owing to the small sample size and to the restricted range of scores and variation within the sample. However, the estimated reliability of both instruments in the language comprehension measures are acceptable for this sample of children.

d. Relationship of the Language Comprehension Measures to the Other Measures of the Test Battery

Correlation coefficients were calculated between the language comprehension measures, the Language Check, and the productive language measures You Say It and Dímelo Tú. Comparisons were made on the performance of children on these measures only in the preferred language. Figure 3 presents these relationships by each language group.

TABLE 9

Internal Consistencies

Reliabilities

Children Preferring Spanish	alpha	Children Preferring English	alpha
Language Check (16*)	.52	Language Check (16)	.44
Dímelo Tú		You Say It	
Part I (27)	.67	Part I (33)	.79
Part IIa	.44	Part II	.47
Part IIb	.18	Part IIb	.07
Part IIIb	.73	Part IIIb	.77
Part IIIc	.66	Part IIIc	.59
Escuchen Este Cuento (25)	.80	Listen to the Story (25)	.86
Listen to the Story (25)	.75	Escuchen Este Cuento (25)	.74
PSI-English (30)	.68	PSI-English (30)	.85
PSI-Spanish (30)	.79	PSI-Spanish (30)	.80

*Numbers in parentheses indicate the number of items in each test.

TABLE 10

Reliability Estimates for Escuchen Este Cuento and Listen to the Story, by Language Group

Spanish Language Group	alpha	N
Escuchen Este Cuento	.75	37
Simple Comprehension*	.62	
Inference and Interpretation**	.64	
Listen to the Story	.85	37
English Language Group		
Escuchen Este Cuento	.87	49
Simple Comprehension	.61	
Inference and Interpretation	.64	
Listen to the Story	.74	49
Simple Comprehension	.60	
Inference and Interpretation	.57	
English and Spanish Language Groups		
Escuchen Este Cuento	.86	88
Simple Comprehension	.79	
Inference and Interpretation	.72	
Listen to the Story	.84	88
Simple Comprehension	.73	
Inference and Interpretation	.72	

*Total number of items in test = 25.

**Number of items in subtest = 15.

***Number of items in subtest = 10.

For the children classified as English preferring, performance on Listen to the Story is related to the performance on the Language Check (r = .51), with You Say It, Part I (r = .62), and with the total score for You Say It (r = .59). These tests appear to be measuring similar language skills. For children classified as Spanish preferring, performance on Escuchen Este Cuento related well to the Spanish version of the Language Check (r = .40), and to a lesser extent to Part I of Dímelo Tú (r = .36). The apparently stronger relationship between the instruments presented to the English preferring children for each of the language measures suggests that these instruments of the test battery differ depending on the language group to which they are being administered. The differences between versions of the instruments of the battery may raise problems in interpreting data from these tests across the two language groups.

Listen to the Story was not completed by 17 children across all four test sites, and a total of 9 children failed to complete Escuchen Este Cuento. As 11 of the 17 cases emerged at Rio Grande where Spanish speaking children predominated and 8 of the cases occurred at the New York site where a number of the children had no knowledge of Spanish, it is assumed that second language testing proved inappropriate for completely monolingual children.

4. Language Production -- Dímelo Tú and You Say It

Dímelo Tú and You Say It were developed by the Educational Testing Service to assess the child's productive language in

Spanish and English by means of several tasks which elicit responses in functional, descriptive, and narrative language. Either Dímelo Tú or You Say It is administered depending on the child's preferred language. For each task the examiner engages in a dialogue with the child for which responses are recorded verbatim. The test is designed to be administered individually. The overall plan of Dímelo Tú closely resembles that of the English version, You Say It, in terms of format, item types, and scoring categories. However, Dímelo Tú is not a direct translation of You Say It and the two measures of productive language do not have all items in common. The tests, nevertheless, are designed to sample similar behavior.

Dímelo Tú and You Say It are divided into three main parts. Part I, Functional Language, of Dímelo Tú has 27 items, while Part I of You Say It has 33 items. In both, the child is shown a series of pictures which require responses employing morphemes such as verbal inflections, plurals, possessives, and adjectival agreement. In Part II, Descriptive Language, for both Dímelo Tú and You Say It, the child is asked to describe some common objects: a button, a nickel, and a penny for Dímelo Tú, and a pencil and shoelaces for You Say It. The examiner asks the child seven questions which are designed to elicit certain properties such as the object's name, color, function, and size, as well as the name of another object that can be used for the same purpose. The examiner records and codes the child's responses.

In Part III of Dímelo Tú and You Say It, the child is presented with a picture (a classroom scene for Dímelo Tú and a

front yard scene for You Say It) and is asked to tell a story about the picture. The child's story is then recorded verbatim by the examiner. For the pilot study the examiner also tape-recorded the child's responses. All taped responses were coded and scored in Juárez and Associates' offices. The child was given three minutes to complete his or her story and then asked six follow-up questions in order to elicit additional information regarding his/her use of various parts of speech.

What follows is a discussion of the results from the pilot study for both Dímelo Tú and You Say It. Parts I, II, and III will be discussed separately. Total test scores and summary data will be presented first, then item and reliability data, followed by a short discussion of each test.

a. Summary Data -- Part I

Table 11 presents the mean test scores and standard deviations for Part I of Dímelo Tú and You Say It. There were no differences resulting from test order, sex, or site for either test. The mean test score for Dímelo Tú across all four sites was 12.8, or 40% of the answers correct, with a standard deviation of 3.81. The mean scores by site ranged from 13.5 at Rio Grande to 9.5 at Corpus Christi.

On You Say It, no differences were observed to result from test order or sex; differences between sites will be discussed later. The mean score was 14.59, or 44% of the items correct, with a standard deviation of 4.9. The mean scores ranged from 18.3 at Milwaukee to 12.0 for both Rio Grande and New York.

TABLE 11

Summary Data for You Say It and Dímelo Tú

You Say It: Part I

Site	Mean*	SD	N
Corpus Christi	14.90	4.50	20
Rio Grande	12.00	0.00	1
Milwaukee	18.27	4.50	11
New York	12.00	5.60	17
Total	14.59	4.90	49
ETS Norm Sample	9.15	7.96	224

Dímelo Tú: Part I

Site	Mean**	SD	N
Corpus Christi	9.50	4.20	4
Rio Grande	13.50	3.50	24
Milwaukee	12.47	5.00	7
New York	12.50	2.12	2
Total	12.84	3.81	37
ETS Norm Sample	14.60	5.02	345

*Number of items = 33.

**Number of items = 27.

Children at the Milwaukee site did significantly better on this test than children at any of the other sites.

The mean scores and standard deviations obtained from the pilot study were compared with those published by the Educational Testing Service (see Table 11). The achievement of the pilot study children was similar to that of lower socioeconomic status kindergarteners of the normed sample. The mean scores for this latter group were 12.8 for You Say It and 14.6 for Dímelo Tú.

The performance of the pilot study children on Dímelo Tú was slightly below that of the pre-kindergarteners in the normed sample, with mean scores of 14.60 for the normed sample and 12.8 for the pilot study sample.

In summary, Dímelo Tú and You Say It, Part I, were rather difficult tests for the children in the pilot study, with Dímelo Tú the more difficult of the two language production measures.

b. Item Analysis Data -- Part I

Table 12 presents item analysis data for Part I of Dímelo Tú and You Say It. Owing to differences in the number of children tested at each of the four sites (i.e., only one child was categorized as preferring English in Rio Grande), item analyses were performed by language group rather than by site.

Levels of item difficulty were calculated for each test. The percentage of children answering the item correctly was used as an index of how difficult the item was. For Dímelo Tú, Part I, the easiest item was number 27, which was correctly answered by 91% of the children. The most difficult item was number 17, with

TABLE 12

Item Data for Dímelo Tú and You Say It

Dímelo Tú Items	Proportion Answering Correctly	r_{bis}
DT 01	0.54	-0.14
DT 02	0.73	0.10
DT 03	0.41	0.17
DT 04	0.13	0.13
DT 05	0.43	0.24
DT 06	0.40	0.05
DT 07	0.49	0.39
DT 08	0.70	0.43
DT 09	0.22	0.04
DT 10	0.19	0.18
DT 11	0.13	0.25
DT 12	0.54	0.34
DT 13	0.49	0.29
DT 14	0.59	0.17
DT 15	0.78	-0.09
DT 16	0.84	0.22
DT 17	0.03	0.26
DT 18	0.81	0.27
DT 19	0.19	0.18
DT 20	0.40	0.50
DT 21	0.35	0.51
DT 22	0.46	0.49
DT 23	0.43	0.52
DT 24	0.43	-0.09
DT 25	0.46	0.33
DT 26	0.78	0.13
DT 27	0.92	0.13

You Say It	Proportion Answering Correctly	r_{bis}
YS 01	0.48	0.31
YS 02	0.50	0.55
YS 03	0.44	0.34
YS 04	0.79	0.40
YS 05	0.77	0.37
YS 06	0.78	0.40
YS 07	0.58	0.03
YS 08	0.50	0.27
YS 09	0.48	0.29
YS 10	0.44	0.32
YS 11	0.31	0.22
YS 12	0.42	0.11
YS 13	0.31	0.05
YS 14	0.19	0.16
YS 15	0.52	0.08
YS 16	0.14	0.40
YS 17	0.62	0.33
YS 18	0.75	0.31
YS 19	0.35	0.29
YS 20	0.50	0.27
YS 21	0.44	0.28
YS 22	0.89	0.30
YS 23	0.29	0.21
YS 24	0.46	0.37
YS 25	0.50	0.24
YS 26	0.77	0.54
YS 27	0.35	0.36
YS 28	0.17	0.35
YS 29	0.31	0.33
YS 30	0.12	0.28
YS 31	0.35	0.46
YS 32	0.17	0.22
YS 33	0.02	0.14

only 2% of the children giving the correct answer. Six items -- 4, 9, 10, 11, 17, and 19 -- were answered correctly by less than 20% of the children.

For You Say It, Part I, item difficulty ranged from the easiest item, number 22, which was answered correctly by 89% of the children, to the most difficult item, number 33, with only 2% of the children responding correctly. Only five items -- 4, 5, 18, 26, and 29 -- were answered correctly by more than 75% of the children. Six items were answered correctly by less than 20% of the children: 14, 16, 28, 30, 32, and 33.

The functional language sections of Dímelo Tú and You Say It were very difficult for the children in the pilot study, with You Say It being slightly easier than Dímelo Tú.

An estimate of the discriminating power of a test is the point biserial correlation, which provides an index of the extent to which an item on a test can predict performance on the total test. The point biserial correlations were calculated for both Dímelo Tú and You Say It. As can be seen from Table 12, the point biserial correlation for each subtest is much lower than the acceptable point biserial correlations for achievement tests in general. Achievement tests assess a broad field of skills and knowledge; therefore, reliability coefficients for this type of instrument are rather modest (r_{bis} = .40). Part I of Dímelo Tú and You Say It may be said to have unacceptable discriminating power.

c. Reliability Estimates -- Part I

The reliability of any set of measurements is logically defined as the proportion of the variance that is due to differences in the ability of the children. Reliability is dependent upon the characteristics of the sample as well as on those of the instrument itself. Reliability coefficients were generated for each language group. Reliability estimates using internal consistency (coefficient alpha) were correlated for both Dímelo Tú and You Say It for the total sample. The alpha coefficient = .67 for Dímelo Tú and .79 for You Say It.

Table 13 presents the estimate of reliability obtained in the publisher's norming study and in the pilot study. As can be seen, estimates from the publisher's study are much higher than those obtained from the pilot study. Some important differences exist between the children who participated in the pilot study and the children in the norm sample. For instance, the publishers included three different socioeconomic status groups, whereas the pilot study sample consisted only of children in the lower socioeconomic status group. The range of ages in the publisher's sample was from 4 to 6 years (48 months through 72 months), while the pilot study ages ranged from 3 years 7 months to 5 years 11 months. Given the slightly different age range and the restricted socioeconomic status variable of our pilot study and the smaller range of test scores, one would expect the pilot study reliability coefficients to be more modest than those obtained in the publisher's sample.

TABLE 13

Reliability Estimates for You Say It*

Score	Items	J&A	PK	K	1st Gr.
Functional Language	33	.79	.94	.94	.92
Description	8	.27**	.69	.67	.59
Narration: Quality	12	.77	.77	.77	.66
Follow-up	8	.59	.78	.75	.65

Reliability Estimates for Dímelo Tú*

Score	Items	J&A	PK	K	1st Gr.
Functional Language	27	.67	.79	.81	.81
Description	7	.31	.48	.50	.54
Narration: Quality	12	.73	.71	.71	.68
Follow-up	8	.66	.57	.62	.59

*Publisher's data compared to data from Juárez and Associates' (J&A) pilot study.

**Part a and b of description were combined to compute reliability coefficient

d. Relationship of Dímelo Tú and You Say It, Part I, to the Other Measures in the Test Battery

The correlation of Dímelo Tú and You Say It with the other language measures in the test battery (see Figure 3) is moderate to high. For Dímelo Tú, Part I (functional language), and the Spanish Language Check, r = .54. For You Say It, Part I (functional language), and the English Language Check, r = .60. Parts II and III of Dímelo Tú and You Say It did not correlate highly in either language edition with the Language Check.

The correlation coefficient for Dímelo Tú with the language comprehension measure Escuchen Este Cuento showed that these two tests were not related. For You Say It, the functional language section and the total test score correlated moderately highly with Listen to the Story. The correlation coefficient between Part I and Listen to the Story was .62, while r = .59 for the total test score and Listen to the Story. Thus it appears that the functional language sections of You Say It and Listen to the Story are tapping the same ability in the children tested in English. But this relationship did not appear for the Spanish speaking children in the pilot study.

e. Part IIa

The mean score for the total sample for Dímelo Tú, Part IIa, was 2.9 out of a possible 7 items, with a standard deviation of .97. The mean scores ranged from 3.29 at Rio Grande to 2.14 at Milwaukee.

The mean score for the total sample represented slightly less than 50% of the items answered correctly (Table 14). The most difficult item was number 4, which asked the child to describe the function of a button: no one answered this item correctly. The easiest item was item 1, which asked the child to label the button, with 84% of the children responding correctly.

The discriminating power of the 7 items making up Part IIa was reflected in the point biserial correlation, $r_{bis} = .23$. This section has adequate discriminating power, but its reliability with an alpha = .44 is rather tenuous, given that it contains so few items. Part IIa correlates moderately well with the total test score on Dímelo Tú and with the Spanish edition of the PSI.

The mean score for the total sample on You Say It, Part IIa, was 3.53 out of a possible 6. By site the highest mean score of 3.85 was obtained at Corpus Christi, while the lowest of 2.0 occurred at Rio Grande. The item difficulties for You Say It, Part IIa, are presented in Table 15. The easiest item was number 1, with 95.9% of the children answering correctly. The most difficult item was number 4, with only 14% of the children scoring a correct response. Item 4 asked the child to give the function of a pencil. Overall, this section proved to be moderately difficult for the children taking this test.

The reliability of this instrument is represented by alpha = .47. However, owing to the small number of items in this subtest, this estimate cannot be considered an accurate one. This subtest possesses acceptable discriminating power.

TABLE 14

Item Analysis Data for Dímelo Tú

Response Mode for Part IIa - Description

Items	Correct Responses in English	%	Correct Responses in Spanish	%	Incorrect Responses	%	No Response	%
1	1	2.7	30	81.1	6	16.2	0	0
2	14	37.8	6	16.2	15	40.5	2	5.4
3	0	0	28	75.7	9	24.3	0	0
4	0	0	0	0	34	91.9	3	8.1
5	5	13.5	8	21.6	23	62.2	1	2.7
6	0	0	29	78.4	7	18.9	1	2.7
7	0	0	6	16.2	16	43.2	15	40.5

TABLE 15

Item Analysis Data for You Say It

Response Mode for Part IIa - Description

Items	Correct Responses in English	%	Correct Responses in Spanish	%	Incorrect Responses	%	No Response	%
1	47	95.9	0	0	2	4.1	0	0
2	31	63.3	0	0	14	28.6	4	8.2
3	45	91.8	0	0	4	8.2	0	0
4	7	14.3	0	0	36	73.5	6	12.2
5.	30	61.2	0	0	15	30.6	4	8.2
6	13	26.5	0	0	26	53.1	10	20.4

The descriptive language part of Dímelo Tú and You Say It (Part IIa) was of average difficulty for each of the children taking the test in the pilot study. This test provides information on the child's use of descriptive language within the context of describing a common object. Discriminating power varies from item to item. The most difficult on both tests is item 4, which asks the children to give the function of the object. In general, the discriminating power of this subtest is adequate. In addition, there is a fairly high correlation between this subtest and other measures on the test battery. This is especially apparent in the relationship between the subtest and both the Preschool Inventory and Listen to the Story.

f. Part IIb

For Part IIb of both You Say It and Dímelo Tú, the examiner codes the response in accordance with a pre-established set of categories. The examiners in the pilot study not only recorded the child's responses to these items by writing them down verbatim, but also taped the responses so that they could be later coded at the Juárez and Associates offices in Los Angeles. Since coders in the Juárez and Associates offices utilized a pre-classified set of categories to code the child's responses on this test, the item analysis is slightly different here. For each of the categories, item difficulty was defined as the percentage of students in the pilot study who were able to give a response that fit within each pre-classified category. These categories were labeling

of the object, indicating the function of the object, indicating the color of the object, specifying the shape of the object or materials that the object is made of, assigning a number value to the object (i.e., a nickel is equal to five pennies), indicating other physical characteristics of the object, and finally making a comparative statement about the characteristics of the object, such as, "It is as big as a ball" or "as small as a needle." Since the coding of these answers requires judgments about qualitative features of the object, the percentage of children responding to each of the categories was tabulated.

(1) Item Analysis

Part IIb of Dímelo Tú requests that the child describe two shoelaces and respond to questions about them. The item analysis (Table 16) shows that, other than correctly labeling the shoestrings in Spanish (item 1) and describing their function in Spanish (item 5), the majority of the children found the inquiries too difficult. More than half the children could not describe the color of the shoestrings, and a number of "no responses" were associated with the more difficult items.

Part IIb of You Say It featured questions about a penny and a nickel. As can be seen from Table 17, item 1, which involved labeling the object, was completed by 83.7% of the sample, and item 6, which involved indicating the use of the object, was correctly answered by 55.1% of the sample. The remainder of the items in this part of You Say It proved quite difficult for the

TABLE 16

Item Analysis Data for Dímelo Tú

Response Mode for Part IIb - Description

Item	Correct Responses in English	%	Correct Responses in Spanish	%	Incorrect Responses	%	No Response	%	Categories
1	1	2.7	29	78.4	5	13.5	2	5.4	Specific Label
2	3	8.1	6	16.2	21	56.8	7	18.9	Color
3	0	0	C	0	29	78.4	8	21.6	Shape
4	0	0	0	0	29	78.4	8	21.6	Material
5	0	0	20	54.1	13	35.1	4	10.8	Functional
6	0	0	1	2.7	28	75.7	8	21.6	Number of Value
7	0	0	3	8.1	26	70.3	8	21.6	Other Physical Characteristic
8	0.	0	0	0	29	78.4	8	21.6	Comparative Characteristic

TABLE 17

Item Analysis Data for You Say It

Response Mode for Part IIb - Description

Item	Correct Responses in English	%	Correct Responses in Spanish	%	Incorrect Responses	%	No Response	%	Categories
1	41	83.7	0	0	7	14.3	1	2.0	Specific Label
2	13	26.5	2	4.1	31	63.3	3	6.1	Class
3	1	2.0	0	0	43	87.8	5	10.2	Color
4	0	0	0	0	44	89.8	5	10.2	Shape
5	1	2.0	0	0	43	87.8	5	10.2	Material
6	27	55.1	1	2.0	17	34.7	4	8,2	Function
7	4	8.2	0	0	40	81.6	5	10.2	Number of Value
8	10	20.4	0	0	35	71.4	4	8.2	Other Physical Characteristics
9	2	4.1	0	0	42	85.7	5	10.2	Comparative Characteristics

children, as a majority in each instance failed to note the class, color, shape, material, number, value, or other physical or comparative characteristics of the object. The categorization used in the scoring was clearly incompatible with the children's ability or experience, as most of the children in this sample failed to perform as the examination scoring schedule anticipated.

(2) Reliability Estimates

As before, the correlation coefficient was computed by scoring as correct the acceptable responses to specific categories and as incorrect the responses that did not fit the specific categories or those occasions when the child refused to answer.

The correlation between the children's performance in Part IIb of Dímelo Tú with the total test score on Dímelo Tú was calculated by assigning correct values to each of the questions asked in the test: if the child gave a specific word that corresponded to the categories that question was intended to probe, that response was coded as correct. If the child was unable to give an appropriate response, or did not reply to the question, those responses were coded as incorrect. The reliability coefficient for Dímelo Tú, Part II (internal consistency), was alpha = .18. The children were able to give appropriate responses to questions that probed for specific labels 78.4% of the time. The children were also able to give functional descriptions on 54% of the appropriate occasions. The children in the pilot study were not able to give responses having to do with other physical or comparative characteristics of the objects.

The reliability coefficient for You Say It was alpha = .07. The children were able to give appropriate responses to the question that probed for a specific label 83.7% of the time, whereas for the nickel the responses were correct only 55.1% of the time. The children in the pilot study were generally not able to respond appropriately to items requiring them to provide the objects' color, shape, material, number value, or comparative characteristics.

Both Dímelo Tú and You Say It, Part IIb, were difficult for these children. The children's test behavior on Part IIb would indicate that this test was tapping information that was unfamiliar to the children. This is particularly important given that children in the pilot study had about six months of Head Start experience.

g. Part IIIb

(1) Dímelo Tú

In Part IIIb of Dímelo Tú (narration), the child is presented with a picture about which s/he must tell a story in Spanish. In scoring the test, a predetermined set of characteristics was used. Table 18 shows that most children were able to label at least four objects in the picture in response to item 1, use a verb (item 2), and use a complete sentence (item 5). The remainder of the anticipated morphological content of the story was difficult for the children, as most of the sample failed to use an adjective (item 3), an adverb (item 4), or an infinitive (item 6), or to

TABLE 18

Item Analysis Data for Dímelo Tú

Response Mode for Part IIIb - Narration

Items	Correct Responses in English	%	Correct Responses in Spanish	%	Incorrect Responses	%	No Response	%	Categories
1	0	0	30	81.1	6	16.2	1	2.7	Labeling
2	0	0	35	94.6	1	2.7	1	2.7	Verb
3	3	8.1	18	48.6	15	40.5	1	2.7	Adjective
4	0	0	21	56.8	15	40.5	1	2.7	Adverb
5	0	0	33	89.2	3	8.1	1	2.7	Syntax
6	0	0	4	10.8	32	86.5	1	2.7	Verbal
7	0	0	28	75.7	8	21.6	1	2.7	Spatial
8	0	0	6	16.2	30	81.1	1	2.7	Plot Extension
9	0	0	7	18.9	29	78.4	1	2.7	Sequence
10	1	2.7	18	48.6	17	45.9	1	2.7	Character
11	0	0	8	21.6	28	75.7	1	2.7	Emotion
12	0	0	1	2.7	35	94.6	1	2.7	Comparison

extend the plot (item 8), sequence events (item 9), extend characters (item 10), use an emotional word (item 11), or make one comparison (item 12). These latter items may not appear in children's descriptions because of the children's lack of ability and experience relative to the tasks. Therefore, some of the precoded categories were apparently inappropriate for the sample tested.

(2) You Say It

In this part the examiner shows a picture to the child as in Dímelo Tú. The child's story is then coded in accordance with the pre-established set of performance criteria and standards. Items 1, 2, 5, and 7 (concerning labeling, use of a verb, use of a complete sentence, and spatial descriptions, respectively) were answered in English by a high percentage of children (Table 19). Items 6, 8, 9, 11, and 12 (concerning use of a verb, plot extension, sequence of events, use of an emotional word, and comparisons, respectively) reflected the inability of a large percentage of children to respond correctly. Items 3, 4, and 10 (use of an adjective, adverb, and character extension) also proved to be difficult for the sample with approximately 40% to 50% of the children responding correctly to these items.

h. Part IIIc

Part IIIc of Dímelo Tú involves follow-up questions to the picture presented in Part IIIb. The examiner asks questions of

TABLE 19

Item Analysis Data for You Say It

Response Mode for Part IIIb - Narration

Items	Correct Responses in English	%	Correct Responses in Spanish	%	Incorrect Responses	%	No Response	%	Categories
1	37	75.5	0	0	9	18.4	3	6.1	Labeling
2	43	87.8	0	0	3	6.1	3	6.1	Verb
3	21	42.9	0	0	25	51.0	3	6.1	Adjective
4	19	38.8	0	0	27	55.1	3	6.1	Adverb
5	41	83.7	0	0	5	10.2	3	6.1	Syntax
6	17	34.7	0	0	29	59.2	3	6.1	Verbal
7	23	46.9	0	0	23	46.9	3	6.1	Spatial
8	16	32.7	0	0	30	61.2	3	6.1	Plot Extension
9	6	12.2	0	0	40	81.6	3	6.1	Sequence
10	18	36.7	0	0	28	57.1	3	6.1	Character Extension
11	2	4.1	0	0	44	89.8	3	6.1	Emotion
12	0	0	0	0	46	93.9	3	6.1	Comparison

the children regarding the picture used for the narration. Table 20 shows that most children were able to answer correctly in Spanish item 1 (label object), item 2 (use a verb), and item 4 (sequence event), while fewer, but still a majority, were able to describe the position of a girl in the picture (item 6). A slight majority of the sample did not use an emotional word (item 5), did not use a complete sentence (item 7), or were unable to extend the characters displayed in the picture (item 8). The children also had difficulty in using the modifiers (item 3), as evidenced by the number of incorrect responses and nonresponses to this item.

In Part IIIc of You Say It, the child is asked to respond to particular questions regarding a picture presented by the examiner. The responses are then coded in accordance with a pre-established set of categories. As can be seen from Table 21, most children were able to label objects (item 1), use verbs (item 2), and extend characters (item 8). At least 50% of the children were able to respond correctly to many of the other questions. Items 3 and 7 had high incorrect response rates, however, as children failed to use modifiers or employ a complete sentence. As is evidenced by the low response rates, the parts of speech categories used in this section to evaluate the child's responses appear to be inappropriate for these children.

In conclusion, the measures of language production, Dímelo Tú and You Say It, are rather difficult and long for the children tested in the pilot study. Many of the questions in these measures,

TABLE 20

Item Analysis Data for Dímelo Tú

Response Mode for Part IIIc - Follow-up Questions Narration

Items	Correct Responses in English	%	Correct Responses in Spanish	%	Incorrect Responses	%	No Response	%	Categories
1	1	2.7	31	83.8	5	13.5	0	0	Labeling
2	0	0	32	86.5	4	10.8	0	0	Verb
3	0	0	14	37.8	16	43.2	6	16.2	Modifier
4	0	0	26	70.3	8	21.6	2	5.4	Sequence
5	0	0	13	35.1	19	51.4	4	10.8	Emotion
6	0	0	22	59.5	11	29.7	3	8.1	Spatial
7	0	0	15	40.5	21	56.8	0	0	Syntax
8	1	2.7	16	43.2	19	51.4	0	0	Character Extension

TABLE 21

Item Analysis Data for You Say It

Response Mode for Part IIIc - Follow-up Questions Narration

Items	Correct Responses in English	%	Correct Responses in Spanish	%	Incorrect Responses	%	No Response	%	Categories
1	43	87.8	0	0	5	10.2	1	2.0	Labeling
2	47	95.9	0	0	1	2.0	1	2.0	Verb
3	22	44.9	0	0	23	46.9	4	8.2	Modifier
4	29	59.2	0	0	17	34.7	3	6.1	Sequence
5	31	63.3	0	0	15	30.6	3	6.1	Emotion
6	32	65.3	2	4.1	13	26.5	2	4.1	Spatial
7	26	53.1	0	0	22	44.9	1	2.0	Syntax
8	35	71.4	0	0	13	26.5	1	2.0	Character Extension

particularly those in the descriptive and narrative language use sections (Parts IIa, IIb, IIIb, and IIIc), appear to be inappropriate for this specific population of children. Reliability of coefficients for these subtests range from inadequate to acceptable (r = .18 to r = .73). The low reliability coefficients are due in part to the fact that the test as a whole attempts to assess a variety of skills and abilities, thus limiting the number of items that can be utilized to test specific areas of knowledge.

Although testers scored a "no response" on all items for Dímelo Tú and You Say It when the child did not respond to a question, there was no indication that the tests were not completed due to item difficulty.

5. Language Production -- Bilingual Syntax Measure

The Bilingual Syntax Measure (BSM), developed by Burt, Dulay, and Hernandez (1975), is used to assess the oral proficiency of children in their nonpreferred language. The test has been field tested previously with Mexican American, Anglo, Puerto Rican, and Cuban children and has been used in other national evaluations where it received positive reviews from a number of specialists in bilingual assessment.

The test requires only a short time to administer and is given individually. The administration follows an informal conversational pattern, asking questions about several pleasant pictures. The examiner helps the child to feel at ease by presenting

pictures and asking preliminary questions, then tests the child on twenty-five questions with appropriate pictures and recording responses verbatim in an answer booklet.

The initial scoring procedure with the Bilingual Syntax Measure is to divide the children into two general categories: children whose responses are not scored, and children whose responses are scored. Within these divisions five levels of proficiency are included, with the first two falling under division one and the last three under division two. Level one children were those who answered fewer than three out of five initial questions in any language, while level two comprised those children who answered fewer than thirteen of the BSM questions, answered all of the questions but in a language other than that in which the test was administered, or had thirteen or more unscorable responses. Level three, four, and five children were those who had a Syntax Acquisition Index (SAI) of between 45 and 84, 85 and 94, and 100, respectively. This index was calculated by dividing a child's total attempts to provide answers (Child's Response Value) by his/her score in reaching syntactical objectives (Development Form Value).

This computation yielded a language development level, or SAI, for each child whose responses were scored. The nature of the BSM required a different approach to the examination of its technical adequacy than the other measures of the battery. The test addresses oral language proficiency across a number of syntactical domains. Therefore, summary data consist of a frequency

count of second language proficiency level by site, while item analysis is limited to an investigation of the variation among the mean SAI scores by items across sites. The discussion of consistency deals with comparability of scoring methods.

a. Summary Data

The results of the BSM administration by level, language, and site is presented in Table 22. The table indicates some 58%, or 52 of the 90 children for whom scores are available, scored at level one. This indicates that the majority of the sample is at the very beginning of the process of learning a second language. Some 16% of the total sample scored at level two, indicating that these children are able to understand varying amounts of everyday language. In other words, they may be able to produce some short sentences or questions but are unable to use the secondary language to communicate thoughts or opinions. Twelve percent of the sample scored at level three, indicating that the respondents could make themselves understood but are still experiencing some difficulties with the secondary language. For example, they may omit nouns and verbs or misform such words, or they may use pronouns incorrectly. Of the total sample, 13% (12 children) scored at level four. These individuals have little difficulty in communicating their ideas in the secondary language. They control a number of basic grammatical structures, including some verb endings and pronoun forms. Only one child among all the sites scored at level five, thereby demonstrating native or near native

TABLE 22

BSM Results by Language, Level, and Site

LEVEL	CORPUS CHRISTI		NEW YORK		RIO GRANDE		MILWAUKEE		TOTAL	
	ENGLISH	SPANISH	ENGLISH	SPANISH	SPANISH	ENGLISH	ENGLISH	SPANISH	ENGLISH	SPANISH
I	5	2	9	1	--	22	8	5	22	30
II	5	--	4	2	1	1	1	--	11	3
III	5	--	2	1	--	--	2	1	9	2
IV	6	--	1	2	--	1	2	--	9	3
V	--	1	--	--	--	--	--	--	--	1
TOTAL	21	3	16	6	1	24	13	6	51	39

proficiency in the secondary language. As these 13 children may be considered bilingual, the BSM can be used during data analysis to establish more precise language categories than those used for purposes of testing.

The distribution across levels varies by sites, with Corpus Christi demonstrating a rather uniform distribution among those who took the test in English. The Milwaukee and New York samples also reflect some distribution across levels but do not demonstrate the balanced distribution of Corpus Christi. The most skewed distribution of a site sample is that of Rio Grande, where 22 of the 25 children who took the test scored at level one.

b. Item Analysis

In order to obtain more specific information, BSM responses were subjected to an item analysis. Response mode was noted (i.e., whether responses were in English or Spanish) and the mean Syntax Acquisition Index (SAI) score was intended as a measure of saliency and level of difficulty for each item, by site, for each of the two language groups. The data presented in Table 23 are based on the 22 children from all sites who scored above level two on the five-level scale and hence were usable for an item analysis. Ten children of this higher scoring group took the test in Spanish at Corpus Christi, while the remaining 12 children were distributed across the other sites in both languages. The impact of this small sample size and the distribution across sites for an item analysis is apparent in Table 23. Only 2

BSM - Syntax Acquisition Index (SAI)

Item Analysis Mean Frequency Distribution Per Site

SAI - ENGLISH RESPONSE

Item	Corpus Christi	Rio Grande	Milwaukee	New York
1	100	100	50	66
2*	*	*	*	*
3	42	0	50	60
4	100	75	80	27
5	100	50	33	66
6	75	100	75	43
7	50	0	25	57
8	75	100	50	50
9	50	100	50	33
10*	*	*	*	*
11	100	0	55	60
12*	*	*	*	*
13	57	0	50	25
14	50	100	61	97
15	50	60	65	63
16	93	100	71	14
17	93	100	100	83
18	100	100	67	80
19	29	40	41	60
20	100	100	83	78
21*	*	*	*	*
22	90	100	69	70
23	33	100	50	19
24	100	67	73	39
25*	*	*	*	*
N =	2	1	2	3

SAI - SPANISH RESPONSE

Corpus Christi	Rio Grande	Milwaukee	New York
85	—	100	100
77	—	100	100
45	—	75	57
66	—	83	62
58	—	87	87
10	—	47	47
73	—	100	60
84	—	100	
*	*	*	*
*	*	*	*
67	—	95	67
*	*	*	*
93	—	44	92
66	—	97	94
*	*	*	*
8	—	38	33
*	*	*	*
77	—	100	97
*	*	*	*
70	—	90	100
40	—	77	62
39	—	64	70
73	—	96	100
22	—	0	48
*	*	*	*
10	0	1	3

N = 22 Number of children

* These items are not scored because they are used to maintain the conversation between the examiner and the child.

children at Corpus Christi, 1 at Rio Grande, 2 at Milwaukee, and 3 at New York took the test in English. As can be seen, there was little consistency among the mean SAI scores by items across sites, a result that is likely due in large part to the small number of children at each site. A similar result can be seen in the sample whose responses reached levels three, four, or five in Spanish at the four sites. While there were 10 such children at Corpus Christi, none at Rio Grande, 1 at Milwaukee, and 3 at New York, the total sample size was insufficient to make any judgments regarding patterns of either saliency or difficulty by item.

c. Methods Utilized for Coding and Scoring

Two methods were utilized for coding and scoring the BSM. The first method followed the directions provided in the BSM Technical Handbook (1975), which focuses on comparing natural language utterances to their idealized forms. The second employed a more simplified method designed to assign children to various language proficiency levels. Both methods were examined for ease of scoring and the quality of the information supplied.

The BSM Technical Handbook outlines a coding and scoring system based on a comparison of the Child's Response Values (CRV; the child's actual responses to test questions) to the Child's Developed Form Value (DFV; the child's ideal responses). These two values, CRV and DFV, are then computed to yield the child's language proficiency level, the Syntax Acquisition Index (SAI). The formula for computing the SAI is $\frac{CRV}{DFV} \times 100 = SAI$. The SAI

provides the basis for assignment to various language proficiency levels. Although this method provides children with a good opportunity to demonstrate language competence by allowing for language switching and a certain amount of grammatical error, its scoring procedure demands a high level of expertise. It also assigns children to broad range levels which tend to mask differences between children.

In additon to the above scoring system, a second procedure was followed. This simplified procedure directly assigns children to various proficiency levels based on their performance.

Although the simplified procedure is very easy to score, it does not make any provision for language switching or minor grammatical errors. Moreover, it makes no allowances for children's attempts to respond to questions. It simply discriminates on the basis of correctness or incorrectness. However, due to practical considerations in scoring, this method may prove to be more useful in easily identifying children's language development levels.

To test the comparability of results utilizing either of the two methods for scoring the BSM, we computed a correlation between scores obtained from one method with scores obtained from the other method. The kappa (k) correlation coefficient obtained, $k = .80$, indicated that these two methods were similar in their distribution of children to the various levels of proficiency. A final determination as to which scoring method is most appropriate will be made prior to the Fall pretest.

The English and Spanish versions of the BSM were not completed by a total of 64 children across all sites. Because this test had the most specific criteria for discontinuing its administration, it was assumed that more children would be judged as unable to complete the test. The higher noncompletion rate, which was generally equal for both language groups, is likely to reflect, therefore, a greater sensitivity to difficulty than the other tests in the battery.

6. Preschool Inventory

The Preschool Inventory (PSI), developed by the Stanford Research Institute (SRI), was designed as a brief assessment and screening device for children from 3 to 6 years old in order to assess their general cognitive skills and concepts. This 30-item version of the test also assesses a number of verbal, quantitative, perceptual, and motor skills. The PSI is generally felt to be highly sensitive to educational intervention and has been used in a number of large-scale evaluations. In addition, it assesses aspects of concept development such as information processing, conceptual inference, understanding of relational terms, and time relationships.

Shortly after the publication of the revised edition of the PSI, researchers in various sections of the United States requested permission from the Educational Testing Service, the owner of the copyright, to translate the PSI into Spanish. The Spanish edition of the PSI shares the two primary goals of the English edition:

(1) to provide educators with a measuring device that will highlight the skills and knowledge that a child from a low socioeconomic status may have and those s/he may lack upon entering school, and (2) to make available an instrument that is sensitive to a child's ongoing experience.

Since the children participating in the pilot study were enrolled in various bilingual/bicultural Head Start programs, each child was administered the PSI in both languages, first in their secondary language and then after two days in their primary language. This order was intended to control for test-wiseness: the child would not be expected to be able to answer as many items in his/her secondary language. An additional analysis (a t-test for differences between means obtained on both editions of the PSI) was performed to look at the similar characteristics of the two language editions of the PSI.

a. Summary Data

Table 24 presents the mean scores and standard deviations for children preferring Spanish on both language editions of the PSI. Differences across sites were obtained when these children took the PSI in their primary language. The children at the Rio Grande site scored the highest, with a mean of 21.6, and children at the Corpus Christi site scored the lowest, with a mean of 16.75. The mean scores across all sites for children preferring Spanish were 20.35 for their primary language and 11.13 for their secondary language.

TABLE 24

Summary Data for Spanish Preferring

Children in Preschool Inventory

Language of Test	Means		S.D.		N
	Spanish	English	Spanish	English	
Sites					
Corpus Christi	16.75	17.00	4.60	4.80	4
Rio Grande	21.60	10.50	6.40	9.00	24
Milwaukee	19.70	9.0	4.19	8.45	7
New York	17.76	13.06	6.32	5.75	2
TOTAL	20.35	11.13	5.81	8.46	37

Number of items = 30

Table 25 presents the mean scores and standard deviations for English speaking children taking both language editions of the PSI. (Owing to the small number of children within each cell, no tests of differences were performed.) The mean scores on the English edition of the PSI obtained by children who preferred English ranged from 24.0 at the Rio Grande site to 17.7 at the New York site. For the same children taking the English edition (in their second language) the mean scores ranged from 17.65 at Corpus Christi to 7.76 at New York. The overall mean score for English speaking children across sites was 20.32 for their preferred language and 12.87 for their secondary language.

In conclusion, children in the pilot study taking the PSI in their primary language on the average answered at least two-thirds of the questions correctly. As expected, children taking the PSI in their secondary language did not perform as well as in their primary language. When the children in our pilot study took the PSI in their primary language, they found the test moderate to easy. When they took it in their secondary language, they found the PSI much more difficult, and were only able to answer approximately one-third of the items correctly. Although the PSI is designed to assess the child's knowledge of concepts, his/her performance on this test is also affected by his/her skill and knowledge of the language in which the test is given.

TABLE 25

Summary Data for English Preferring

Children in Preschool Inventory

Language of Test	Means		S.D.		N
	English	Spanish	English	Spanish	
Sites					
Corpus Christi	21.65	17.65	3.4	4.44	20
Rio Grande	24.00	17.00	0.00	0.00	1
Milwaukee	22.63	11.73	5.37	11.35	11
New York	17.7	7.76	6.31	7.11	17
TOTAL	20.32	12.87	5.07	7.41	49

Number of items = 30

b. Item Analysis Data

Since each child was administered the PSI in both English and Spanish, item analysis data includes analyses of item difficulty and discriminations between the two language editions.

For children preferring Spanish taking the Spanish edition of the PSI, item difficulties ranged from all the children answering item 26 correctly to only 11% answering item 15 correctly, with at least 75% of the children answering items 1, 3, 4, 5, 6, 7, 11, 16, 23, 24, 25, 26, and 29 correctly. The most difficult question once again was "¿En qué dirección se mueve una catarata?" (item 15; see Table 26).

Table 26 presents the item difficulties for children preferring Spanish taking the English edition of the PSI. These ranged from all of the children answering item 26 correctly (item 26 required the child to draw a circle) to only 5% of the children answering items 3 and 14 correctly. (Item 14 asked, "If you wanted to read something what would you do?" Item 3 asked, "Show me your heel.")

Item difficulty for children preferring English on the English version of the PSI ranged from 96% correct on item 25 to 15% on item 3. When English preferring children were tested in Spanish, item difficulty went from 97% on item 26 to 9% on item 15 (see Table 27).

TABLE 26

Item Data: Spanish Preferring Children
in Both Language Versions of the Presschool Inventory

Spanish Items	Proportion Answering Correctly	r_{bis}	N
PSIS 01	0.85	0.22	27.0
PSIS 02	0.48	0.44	27.0
PSIS 03	0.81	0.30	27.0
PSIS 04	0.89	-0.05	27.0
PSIS 05	0.93	-0.32	27.0
PSIS 06	0.93	0.24	27.0
PSIS 07	0.81	0.44	27.0
PSIS 08	0.63	0.33	27.0
PSIS 09	0.74	0.52	27.0
PSIS 10	0.41	0.17	27.0
PSIS 11	0.89	0.43	27.0
PSIS 12	0.59	0.31	27.0
PSIS 13	0.52	0.34	27.0
PSIS 14	0.44	0.52	27.0
PSIS 15	0.11	0.04	27.0
PSIS 16	0.89	0.68	27.0
PSIS 17	0.70	0.57	27.0
PSIS 18	0.37	0.20	27.0
PSIS 19	0.56	0.29	27.0
PSIS 20	0.70	0.51	27.0
PSIS 21	0.74	0.26	27.0
PSIS 22	0.63	-0.02	27.0
PSIS 23	0.96	0.45	27.0
PSIS 24	0.96	0.45	27.0
PSIS 25	0.96	-0.20	27.0
PSIS 26	1.00	0.00	27.0
PSIS 27	0.55	0.27	27.0
PSIS 28	0.70	0.30	27.0
PSIS 29	0.85	0.49	27.0
PSIS 30	0.59	0.58	27.0

English Items	Proportion Answering Correctly	r_{bis}	N
PSIE 01	0.47	0.01	19.0
PSIE 02	0.68	-0.15	19.0
PSIE 03	0.05	0.37	19.0
PSIE 04	0.84	-0.41	19.0
PSIE 05	0.74	0.23	19.0
PSIE 06	0.89	0.56	19.0
PSIE 07	0.68	0.12	19.0
PSIE 08	0.42	0.18	19.0
PSIE 09	0.63	0.73	19.0
PSIE 10	0.53	0.44	19.0
PSIE 11	0.68	0.53	19.0
PSIE 12	0.26	0.07	19.0
PSIE 13	0.26	0.30	19.0
PSIE 14	0.05	0.25	19.0
PSIE 15	0.26	0.33	19.0
PSIE 16	0.68	0.66	19.0
PSIE 17	0.58	0.61	19.0
PSIE 18	0.21	0.20	19.0
PSIE 19	0.21	0.13	19.0
PSIE 20	0.58	0.48	19.0
PSIE 21	0.26	0.10	19.0
PSIE 22	0.68	0.12	19.0
PSIE 23	0.79	-0.10	19.0
PSIE 24	0.42	0.01	19.0
PSIE 25	0.95	0.24	19.0
PSIE 26	1.00	0.00	19.0
PSIE 27	0.58	0.15	19.0
PSIE 28	0.84	0.13	19.0
PSIE 29	0.84	0.17	19.0
PSIE 30	0.89	0.28	19.0

Item Data: English Preferring Children

Performance in Both Versions of the Preschool Inventory

English Items	Proportion Answering Correctly	r_{bis}	N
PSIE 01	0.83	0.44	47.0
PSIE 02	0.53	0.40	47.0
PSIE 03	0.15	0.15	47.0
PSIE 04	0.66	0.43	47.0
PSIE 05	0.85	0.28	47.0
PSIE 06	0.91	0.15	47.0
PSIE 07	0.87	0.29	47.0
PSIE 08	0.25	0.19	47.0
PSIE 09	0.64	0.55	47.0
PSIE 10	0.64	0.63	47.0
PSIE 11	0.74	0.49	47.0
PSIE 12	0.53	0.31	47.0
PSIE 13	0.51	0.55	47.0
PSIE 14	0.45	0.39	47.0
PSIE 15	0.42	0.35	47.0
PSIE 16	0.89	0.24	47.0
PSIE 17	0.87	0.09	47.0
PSIE 18	0.21	0.13	47.0
PSIE 19	0.38	0.32	47.0
PSIE 20	0.85	0.57	47.0
PSIE 21	0.74	0.50	47.0
PSIE 22	0.81	0.34	47.0
PSIE 23	0.91	0.22	47.0
PSIE 24	0.91	0.33	47.0
PSIE 25	0.96	0.12	47.0
PSIE 26	0.94	0.25	47.0
PSIE 27	0.72	0.33	47.0
PSIE 28	0.79	0.59	47.0
PSIE 29	0.76	0.52	47.0
PSIE 30	0.70	0.46	47.0
Spanish Items			
PSIS 01	0.82	0.23	33.0
PSIS 02	0.12	0.17	33.0
PSIS 03	0.21	0.29	33.0
PSIS 04	0.76	0.16	33.0
PSIS 05	0.94	0.16	33.0
PSIS 06	0.91	0.02	33.0
PSIS 07	0.51	0.55	33.0
PSIS 08	0.27	0.26	33.0
PSIS 09	0.63	0.44	33.0
PSIS 10	0.48	0.59	33.0
PSIS 11	0.57	0.29	33.0
PSIS 12	0.54	0.41	33.0
PSIS 13	0.27	0.29	33.0
PSIS 14	0.42	0.32	33.0
PSIS 15	0.09	-0.03	33.0
PSIS 16	0.79	0.35	33.0
PSIS 17	0.48	0.62	33.0
PSIS 18	0.03	-0.18	33.0
PSIS 19	0.30	0.35	33.0
PSIS 20	0.67	0.50	33.0
PSIS 21	0.51	0.44	33.0
PSIS 22	0.58	0.36	33.0
PSIS 23	0.76	0.36	33.0
PSIS 24	0.88	0.26	33.0
PSIS 25	0.94	0.32	33.0
PSIS 26	0.97	0.34	33.0
PSIS 27	0.64	0.09	33.0
PSIS 28	0.85	0.15	33.0
PSIS 29	0.64	0.05	33.0
PSIS 30	0.33	0.46	33.0

c. Reliability Estimates

The reliability of the Spanish and English editions of the PSI was estimated by computing the alpha coefficient (see Table 28). For children categorized as preferring Spanish who took the English edition of the PSI, the alpha coefficient was .67, while it was .79 for the Spanish edition. For English speakers taking the Spanish edition, the alpha coefficient equaled .79, and for the same children taking the English edition it increased to .85.

d. The Relationship of the PSI with the Other Measures of the Test Battery

The correlation matrix in Figure 3 was generated for all measures of the test battery to determine the extent to which performance on the PSI on either the Spanish or the English edition was related to performance on the other measures. With regard to children categorized as preferring Spanish, performance on the Spanish edition of the PSI was moderately correlated with performance on Part IIa of Dímelo Tú ($r = .43$). Part IIa of Dímelo Tú requires children to use labels for such concepts as color, size, and numbers. Thus it appears that the Spanish edition PSI is assessing some of the same concepts as Part IIa of Dímelo Tú. For the children who preferred English, the correlation between performance on the English edition PSI and on You Say It is high, with $r = .70$, while for the English PSI and Listen to the Story, $r = .57$. As with the children preferring Spanish taking the Spanish PSI and Dímelo Tú, performance on the

TABLE 28

Estimates of Reliability for the
Spanish and English Versions of the
Preschool Inventory

Language Preference	PSI-English	PSI-Spanish	N
Spanish	alpha = .67	alpha = .79	37
English	alpha = .79	alpha = .85	48

English PSI is related to performance on You Say It. Finally, the performance of children preferring English on the PSI is moderately highly related to performance on the English Language Check (r = .66). Thus, the two versions of the PSI have some of the characteristics of parallel tests.

e. Inter-rater Agreement

A check on the agreement of examiners in scoring and coding children's responses on this test was performed at the New York Head Start center. Two pairs of the six local examiners at this site participated in the inter-rater check. Each pair of raters was instructed to administer both language versions of the PSI. During the paired observations, one examiner administered and scored the child's responses, thus providing two relatively independent observations of each child's test behavior. Eleven children were included in the inter-rater check.

The results for both language versions of the PSI show a lack of complete agreement between the observers in each dyad. Of the 30 items in the PSI, the raters disagreed on 4 items for the English version (items 3, 4, 8, and 14) and 3 items for the Spanish version of the PSI (items 3, 7, and 13). Item 3, "show me your heel," produced disagreements across both language versions. Disagreements were observed between at least one pair for the following items: English version 1, 2, 12, 15, and 18; Spanish version 2, 4, 5, 11, 12, 22, 29, and 30. The amount of disagreement between raters suggests that modifications in both the training procedure and the scoring guides are needed.

The PSI English test was not completed by 8 children at the Rio Grande site and 3 children at the Milwaukee site. Because these two sites enrolled the largest number of Spanish dominant children, it is assumed that the test proved to be either too difficult or inappropriate for them, as they may have had little or no knowledge of English.

The PSI Spanish test was not completed by 1 child at Rio Grande, 5 children at Milwaukee, and 9 children at New York. It is assumed that the test was either more difficult for English speaking children or that these children had little or no knowledge of Spanish.

C. Fairness

The pilot study permitted an examination of test battery fairness along three dimensions: environmental conditions, appropriateness, and difficulty. The first assesses the environmental conditions of testing or test ecology and concerns the relations between the children, the test conditions, and the test itself. The second component assesses the appropriateness of a test for a given target population. Here the principal concern is with the cultural and linguistic loadings of test items as they relate to the background and experience of the children. The third dimension examines the average difficulty of different components of the tests which had been related to certain domains. The concern is with the relationship between specific test items and the component categories generated from cross-model objectives (see Juárez and Associates, July 1978).

1. Environmental Conditions of Testing

Two instruments were used to assess the conditions under which the pilot study was carried out at each site. The testers were asked to provide a general profile of the children under test conditions by means of an observational rating scale and to furnish feedback as to their impressions of the adequacy of the test conditions and the test battery. Such information, while not related directly to test suitability, can be extremely useful in adapting testing procedures and training examiners.

A Pupil Observation Checklist, which combined and modified certain items of the instruments used by High/Scope Educational Research Foundation (Deloria, 1974) in the National Home Start

Evaluation and by Stanford Research Institute (1976) in the National Day Care Cost Effects Study, measured testers' judgments of the social involvement of children in the testing situations. The checklist focused on the general areas of a child's social and attending behaviors during testing. Using a five-point bipolar code, testers were asked to give their perceptions of each child along the following dimensions: resistive/cooperative, shy/sociable, outgoing/withdrawn, involved/indifferent, defensive/agreeable, active/passive, gives up/keeps trying, quiet/talkative, attentive/inattentive.

A similar scaling format was used for the Evaluation Feedback Checklist. This instrument, employed in three of the four sites, provided an opportunity for the examiner to rate the battery of tests, the procedures utilized, and the training experience, as well as to comment on specific problems encountered during testing after the sessions had been completed. Scale dimensions measured testers' perceptions of adequacy of time allotted for tests, children's behavior during the tests, difficulty of the test battery as determined by the children's behaviors, quality of the test, and usefulness of the training which examiners had undergone in actually administering the battery. Final comments could range freely over any topic a tester felt relevant to successful test administration.

Examiners' evaluation of children's social behavior and their rating of the testing environment provide some indication of the fairness of the test. Although, the technical adequacy data identified a number of items which proved difficult for the children, the testers in all sites found the children to exhibit generally

positive behavioral responses to the tests. They felt that almost all children had adequate time, and that the tests were not too difficult for preschoolers. When rating test difficulty, examiners found the battery to be moderately difficult for 26% of the pilot sample. Difficulty was usually defined as the use of words unfamiliar to a child (such as encima, cascada). These conflicting results between the item analysis data and the general behavior ratings suggest that whereas the children appeared to comprehend the test items, their performance on some items may demonstrate a lack of experience with the behaviors necessary to perform adequately under testing conditions.

Testers generally rated the quality of the tests as defined by ease of administration as adequate to good (in only 9 of the 58 cases was quality rated as less than adequate). In such cases, the problems most often identified were the crowded or noisy conditions in which testing took place. The space allocated to these testing situations ranged from noisy, cramped quarters to large rooms with various human and material distractions. While these locations proved inappropriate to the testing objectives, it is hoped that, especially with the oral language tests, a quiet comfortable place for the administration of the tests can be secured at each site in the future.

2. Appropriateness

The issue of concern regarding fairness in this component is the relationship between the tests and the cultural and linguistic background of the children. In some instances, it is not clear whether a test item was simply too difficult developmentally for the children, the item was confusing or poorly presented, or the

item was inappropriate culturally and linguistically. There were circumstances where the test format was confusing or poorly presented. For example, some children were unable to distinguish between dark colors because the color shades in the symbols were not sufficiently distinct; some pictures and illustrations were more attractive to children, and these illustrations were chosen irrespective of the content of an item; and some children became inattentive when the examiner read a story that was too long. All of these observations were made by examiners on rating sheets concerning examination quality and reflect a concern for misinterpretations or competing yet valid interpretations.

There were other items which proved to be unfair because of cultural and linguistic characteristics. For example, children who were judged Spanish dominant had a tendency to respond to items which called for colors and numbers in English. It is assumed that code switching reflects the children's predominant use of English for these symbols involving colors and numbers. Another problem concerned the examiner's use of an appropriate vocabulary word to reflect the cultural and regional background of the children. Such is the case with shoelaces (e.g., cintas, cordones, lazos), balloons, ice cream, and so on, as children may be unable to comprehend one of several generally acceptable words because of their lack of exposure to them. This is especially troublesome in tests which are cumulative (e.g., Dímelo Tú, description 2A and 2B) where the same vocabulary word may be used throughout, thereby putting the child at a disadvantage for the entire subtest. Examples include shoelaces, button, and pencil. In another example of cultural and linguistic appropriateness, there is a need to ensure that all

practice sessions provide for both Spanish and English language experiences. This is a need for example, in the CIRCO Escuchen Este Cuento examination, as materials are provided for practice only in English. Likewise, some examinations lend themselves to coding in one or another language, but when a child switches language, coding becomes quite difficult. This is the case with the Preschool Inventory, where the Spanish and English formats should be separated.

3. Difficulty

It is unrealistic to think that the test battery has covered every curricular objective within each model. By grounding the test items to the most important cross-model objectives, we have attempted to ensure that the test battery provides a fair sample of curricular content. Table 29 gives the average difficulty measured by percentage of children responding correctly to sets of test items which are related to each component. In most cases over 50% of the sample responded correctly to a set of items. Only the descriptive skills component under the domain of language production exhibited relatively high difficulty, a result which may be explained by the cultural and linguistic inappropriateness discussed earlier. The following section on recommendations encourages several changes in the test battery to attend to these difficulties. In general, however, even though these data must be considered at best a gross indicator, they suggest that the test battery is adequately tapping important aspects of what children are learning.

Percent of Items Correct by Domain Component

Percent of Items Correct by Domain Component

DOMAIN	TEST	PERCENTAGE
Language Production*		
Descriptive Skills	EL CIRCO	
	Dímelo Tú-Part 2 (Botón)	48
	Dímelo Tú-Part 2 (Cinta)	26
Conversational Development	EL CIRCO	
	Dímelo Tú-Part 3 (Narration Qualitative)	49
	Dímelo Tú-Part 3 (Follow-up)	62
Linguistic Markers	EL CIRCO	
	Dímelo Tú-Part I	49
Language Comprehension*		
Word Meanings	EL CIRCO	
	Escuchen Este Cuento (Spanish Dominant)	58
	Escuchen Este Cuento (English Dominant)	48
Determining the Main Idea	Escuchen Este Cuento (Spanish Dominant)	55
	Escuchen Este Cuento (English Dominant)	47
Concept Development*		
Visual Discrimination	PRESCHOOL INVENTORY - Spanish Version	
	Spanish Dominant	78
	English Dominant	65
Seriation	EL CIRCO	
	Escuchen Este Cuento (Spanish Dominant)	60
	Escuchen Este Cuento (English Dominant)	46
Number Concepts	PRESCHOOL INVENTORY - Spanish Version	
	Spanish Dominant	58
	English Dominant	42
Classification/Grouping	EL CIRCO	
	Escuchen Este Cuento (Spanish Dominant)	60
	Escuchen Este Cuento (English Dominant)	46
	PRESCHOOL INVENTORY - Spanish Version	
	Spanish Dominant	61
	English Dominant	64
Time Relationships	EL CIRCO	
	Escuchen Este Cuento (Spanish Dominant)	60
	Escuchen Este Cuento (English Dominant)	46
General Knowledge	PRESCHOOL INVENTORY - Spanish Version	
	Spanish Dominant	57
	English Dominant	58

*Domains

TEST	PERCENTAGE
CIRCUS	
You Say It-Part 2 (Pencil)	57
You Say It-Part 2 (Penny and Nickel)	25
CIRCUS	
You Say It-Part 3 (Narration Qualitative)	44
You Say It-Part 3 (Follow-up)	70
CIRCUS	
You Say It-Part I	46
CIRCUS	
Listen to the Story (English Dominant)	45
Listen to the Story (Spanish Dominant)	44
Listen to the Story (English Dominant)	55
Listen to the Story (Spanish Dominant)	46
PRESCHOOL INVENTORY - English Version	
English Dominant	80
Spanish Dominant	70
CIRCUS	
Listen to the Story (English Dominant)	64
Listen to the Story (Spanish Dominant)	65
PRESCHOOL INVENTORY - English Version	
English Dominant	53
Spanish Dominant	44
CIRCUS	
Listen to the Story (English Dominant)	64
Listen to the Story (Spanish Dominant)	65
PRESCHOOL INVENTORY - English Version	
English Dominant	79
Spanish Dominant	64
CIRCUS	
Listen to the Story (English Dominant)	64
Listen to the Story (Spanish Dominant)	65
PRESCHOOL INVENTORY - English Version	
English Dominant	65
Spanish Dominant	53

IV. Recommendations

This section summarizes the results and analyses of the pilot testing. First, the test battery is discussed and recommendations are made about each of the tests and subtests, taking into consideration the psychometric properties, efficiency, and fairness of each. Second, the order, length, and scoring of the tests are considered. Third, procedures for data collection, including the recruitment, selection and training of local testers, and the supervision and monitoring of testing, are examined. Finally, issues in data processing including the administration, coding, and scoring of tests are addressed. Within each area, revisions to be made for the pre-test stage of the evaluation to be carried out in the Fall of 1979 are discussed.

A. Test Battery

1. Language Dominance

Results of the pilot test indicate that both the Spanish and the English versions of the Language Check were easily administered, and that the children generally performed above the 75% correct responses criterion in the language identified as dominant. In addition, performance on this test related significantly to performance on the other language measures of the battery. (Dímelo Tú, $r = .56$; Escuchen Este Cuento, $r = .40$; You Say It, $r = .59$; Listen to the Story, $r = .51$).

The utility of this test is in the identification of the child's level of language development specific to the CIRCO test battery. It is recommended that the test be retained without modification as an indicator of the likelihood that a child can complete the remaining test battery, especially the subtests, and, together with other data, as a measure of language preference.

2. Language Comprehension

a. Spanish Language Comprehension

Escuchen Este Cuento is intended to assess Spanish language comprehension for both Spanish and English dominant children.

The low mean scores for children both in their identified preferred language and in their nonpreferred language (13.46 and 9.30 respectively out of a possible 25) indicated that the test was very difficult for the sample children. Data on the efficiency of administration supplied by the test supervisors suggested that because of the length, organization, quality of graphics, and difficulty of the test, children became easily fatigued and inattentive. In general, items which required inference and interpretation were more difficult for the children. It could be argued that as the majority of these items occurred near the end of the test, the relatively poor showing was related to tiredness. It is equally possible that these domains (see Table 29 under Fairness) are not specific objectives of the curriculum models and that they are beyond the children's experience. It is recommended therefore that the test be modified by eliminating those items

which require inference and interpretation, thereby reducing the test to 15 items. The items which should be removed are numbers 7, 13, 16, 17, 19, 20, 21, 24, and 25. The remaining questions are simple comprehension items. This action will enable a closer relationship to exit between the curricular objectives and the test items, since the curricula emphasize comprehension rather than inference.

b. English Language Comprehension

Listen to the Story is intended to assess English language comprehension for both Spanish and English dominant children. Results of the test indicate that because of the level of difficulty ($\bar{X}$ = 13.00, English preferring; $\bar{X}$ 7.62, Spanish preferring), length, poor graphics, and poor organization, children become easily fatigued and inattentive. As with the Spanish version, it is recommended therefore that the test be modified by eliminating those items which require inference and interpretation and which are not directly related to curriculum objectives, thereby reducing the test to 15 items. The items which should be removed are numbers 7, 13, 16, 17, 18, 20, 21, 24, and 25. These items are similar in content to the items recommended for removal from the Spanish version of the test in order to maintain comparability.

3. Language Production

a. Spanish Language Production

(1) Dímelo Tú

Dímelo Tú comprises sections dealing with functional language, description (two parts), narration (two parts), and follow-up to the narration. The test proved to be too long and difficult for the children. Furthermore, it was difficult to administer and required a subjective and lengthy scoring process.

(a) Functional Language

The Functional Language section of Dímelo Tú, which is intended to assess Spanish language grammatical knowledge, proved the most difficult part of the instrument for the children tested. On the average, children answered less than half the items correctly ($\overline{X}$ = 12.84). Although the range of item difficulty is large (3% correct to 91% correct), the discriminatory function of more than half of the items is inadequate (r_{bis} > .20). Graphics are poor, administration is difficult, and scoring is time consuming for the entire subtest.

If this test is used, it is recommended that, to reduce the time of administration and the difficulty, all items answered correctly by fewer than 25% of the children be eliminated. These include items 4, 9, 10, 11, 17, and 19. These items are characterized by poor coordination between the pictures and the language used in

questions. The overall recommendation for the subtest however, is to eliminate this Functional Language section and use instead the Bilingual Syntax Measure as the measure for Spanish language (see discussion under BMS).

(b) Description

Part IIa. This section contains seven items which ask the child to describe a button's characteristics. Given the subtest's moderately high correlation with the total test (r = .54) and with the Spanish edition of the PSI (r = .43), it is recommended that the entire subtest be retained, despite the extreme difficulty of one item (4, which concerns additional functions of the button) and some difficulty in administering the test. Because of the low number of correct responses (7%) to item 5 regarding the number of holes in a button, a large button (one inch in diameter, red in color) with clearly visible holes should be used.

Part IIb. This section is intended to elicit a description of shoelaces from children. Other than labeling (item 1) and indicating the function of the shoelaces, (item 5), which had 78.4% and 54.1% correct responses respectively, the children had considerable difficulty (all items below 17% correct) with this subtest. It is recommended that this subtest be eliminated because it duplicates many of the skills needed in Description Part IIa and the test's internal consistency is low (alpha = .18), and because shoelaces do not have high relevance to a child's discourse or to the curricula of the programs.

(c) Narration

Part IIIa. This section provides a total count of different words produced by a child within a three-minute period elicited through the child's description of the home environment. It is recommended that this section remain in the battery, as it provides some indication of a child's relatively spontaneous, open-ended vocabulary.

Part IIIb. This section is intended to provide an indication of a child's grammatical competence and ability to characterize familiar settings, characters, and so on in response to established questions. Despite some fatigue, most of the children answered items related to nouns, verbs, or sentence formation correctly (81%, 94%, and 89%, on items 1, 2, and 5 respectively) but had difficulty with using adjectives (48.6% on item 3), which indicates that the subtest is sensitive to grammatical development. With the elimination of the Functional Language and Description, Part IIb, sections of Dímelo Tú, thereby reducing the children's tiredness during testing, it is recommended that this section remain in the battery as it is colorful and engaging and provides an opportunity for a structured analysis of relatively spontaneous or natural language production by Spanish speaking children.

(d) Follow-up to Narration (Part IIIc)

This section intends to elicit grammatical knowledge, causal relationships and spatial characteristics in response to a structured set of questions as the child describes a neighborhood yard scene. Although generally tired at this stage, children had some successful responses on all items, with a range

from 35.1% to 83.8% correct. It is recommended that, with the removal of the Functional Language and Description, Part IIb, sections, this section be retained, as it provides structured responses to predetermined questions, thereby giving the child an additional opportunity to demonstrate language development competency, an important cross-model objective.

English Language Production

(1) You Say It

You Say It contains sections on functional language, description (two parts), narration (two parts), and follow-up to the narration.

(a) Functional Language

This section contains 33 items and is intended to assess English language grammatical knowledge. It was the most difficult subtest for the children who took it. On the average, children answered less than half of the items correctly ($\bar{X}$ = 14.59). Although the range of item difficulty is large (2% correct to 89% correct), the discriminatory function of a number of the items (6) is inadequate (r_{bis} > .20). As with the Spanish version, graphics are poor and fail to retain children's interest, administration is difficult owing to the multiple tasks involved, and scoring is time consuming for the entire subtest. If this test is used, it is recommended that, to reduce the time of administration and the difficulty, all items answered correctly by fewer than 30% of the children be eliminated. This includes items 14, 16, 23, 28,

30, 32, and 33. These items are difficult and are characterized by poor coordination between the picture and the language used in the questions. The overall recommendation, however, as with the Spanish version is to eliminate this section and use instead the Bilingual Syntax Measure as a measure for English language (see discussion under BSM).

(b) Description

Part IIa. This section contains six items which ask the child to describe a pencil's characteristics. Given the subtest's high correlation with the total test (r = .89), it is recommended that the entire subtest be retained despite the extreme difficulty of one item (number 4, concerning further functions of the pencil) and some difficulty in administering the test. In addition, because of the difficulty experienced by many children in relating to a pencil, it is suggested that a crayon, red in color, be used. The crayon is also assumed to provide a more apparent relationship to the curricula of the various program models.

Part IIb. This section is intended to elicit a description of a penny and a nickel from children. Other than labeling (item 1), and indicating the function of the coins, (item 6), which had 83.7% and 55% correct responses respectively, the children have considerable difficulty (all items below 25% correct) with this subtest. It is recommended that this subtest be eliminated, since it duplicates many of the skills needed in Description Part IIa and the test's internal consistency is low (alpha = .07). Further,

the penny and nickel do not, apparently, have high relevance to a child's discourse or to the curricula of the programs.

(c) Narration

Part IIIa. This section provides a total count of different words produced by a child within a three-minute period elicited through the child's description of a classroom scene. It is recommended that this section remain in the battery, as it provides some indication of a child's spontaneous, open-ended vocabulary.

Part IIIb. This section is intended to provide an indication of a child's grammatical competence and ability to characterize familiar settings, characters, and so on in response to established questions. Despite some fatigue, children found graphics engaging and were generally successful with items related to nouns, verbs, and sentence formation (75.5%, 87.8%, and 83.7%, on items 1, 2, and 5 respectively) but had difficulty with other items (all below 50%), which indicates that the subtest is sensitive to grammatical development. It is recommended that with the elimination of the Functional Language and Description, Part IIb, sections of You Say It that this section remain in the battery, as it provides an opportunity for a structured analysis of spontaneous or natural language production by English speaking children.

(d) Follow-up to Narration (Part IIIc)

This section intends to elicit grammatical knowledge, causal relationships, and spatial characteristics in response to a structured set of questions as the child describes a classroom

(2) Length of Test Sessions

As indicated in Figure 4, the total time for the test sessions should range from 75 to 100 minutes.

(3) Length of Scoring

The length of time for scoring is also indicated in Figure 4. It is estimated that scoring time should range from 100 to 145 minutes.

c. Recruitment, Selection, and Training of Local Testers

(1) Recruitment

It is recommended that ample time be allocated for the recruitment of able testers whose characteristics should include the following:

(a) Familiarity with the local culture and language

(b) Experience, preferably with pre-school children

(c) A minimum of a B.A. degree with some units in child development, psychology, bilingual education, or related areas.

(d) Bilingual, biliterate language competency with special strength in grammar and oral production and comprehension.

(2) Selection

The selection and pre-screening of testers should include the following:

(a) Participation in an informal situation to determine basic English and Spanish language verbal ability.

(b) Administration of the BSM in both Spanish and English to test written language skills, including grammar, punctuation, and content.

(c) Recommendations from acceptable local sources.

FIGURE 4

Test Measures and Testing Sequence

Testing Domains	Session	Spanish Dominant		English Dominant		Estimated Time of Administration	Estimated Time of Scoring
		1st Language	2nd Language	1st Language	2nd Language		
Language Preference	1st. Session	Language Check		Language Check		5 to 10 Minutes	10 to 1 Minutes
First Language Production		Dímelo Tú Parts IIa, IIIa, b, c.		You Say It Parts IIa, IIIa, b, c.		10 Minutes	20 to 3 Minutes
Second Language Comprehension	2nd Session		Listen to the Story (Modified, version of 15 items)		Escuchen Este Cuento (Modified Ver- of 15 items)	10 to 15 Minutes	5 to 10 Minutes
First Language Concept Development		PSI Spanish		PSI English		15 Minutes	10 to 15 Minutes
Second Language Production	3rd Session		BSM-E		BSM-S	5 to 10 Minutes	20 to 2 Minutes
First Language Comprehension		Escuchen Este Cuento (Modified version of 15 items)		Listen to the Story (Modified version of 15 items)		10 to 15 Minutes	5 to 10 Minutes
Second Language Concept Development	4th Session		PSI English		PSI Spanish	15 Minutes	10 to 15 Minutes
First Language Production		BSM-S		BSM-E		5 to 10 Minutes	20 to 2 Minutes
	5th Session Pick-up Session						
TOTAL TIME - - -

(3) Training of Local Testers

(a) It is recommended that the training be extended from two to four days owing to the complexity of administering and scoring the tests.

(b) There should be a total of four sessions, one each day.

(c) The contents and sequence of the training should be as follows:

i. Session I: Introduction to the goals and objectives of the evaluation; Juárez and Associates' and the curriculum developer's role in the evaluation; role of the examiner and expectations of the examiner; issues in testing young children; orientation to the Language Check and EL CIRCO CIRCUS Test measures for the Spanish preferring child; operating instructions for the tape recorder.

ii. Session 2: Oral language production measures including the BSM, Dímelo Tú, and You Say It; language comprehension measures, including Listen to the Story and Escuchen Este Cuento.

iii. Session 3: Use of concept development tests; assembling prop packages for administering tests; general review of prior sessions.

iv. Session 4: Introduction to the use of and the procedures for filling out pupil observation checklists, tester's log, and evaluation feedback rating form.

(4) Procedures

It is intended that each training session will rely upon a combination of training techniques, methods, and applications. First, materials and subjects will be introduced through presentations and discussions. Second, the supervisor/trainers will demonstrate or model the administration of the appropriate tests. Third, participants will have an opportunity to practice under the supervision of trainers the proper administration of the various instruments. Fourth, in all instances considerable hands-on experience and feedback will be provided to the testers, enabling them to perfect their administration, coding, scoring, and reporting skills.

There will be a need to include in the training sessions the examiner's manuals prepared by Juárez and Associates, which include descriptions of each test, instructions for coding and scoring each test, sample copies of each test, pupil observation checklists, tester's logs, and evaluation feedback rating forms. In addition, there will be a need to have on site prop packages, tape recorders, and picture stimuli.

d. Supervision and Monitoring of Testing

(1) Test Supervisors

It is recommended that there be a test supervisor for each site. Each test supervisor will be selected and trained by Juárez and Associates. He/she will hold a B.A. degree in education, psychology, or a related field; be bilingual and biliterate in Spanish and

English; have experience with training testers and in the administration of tests; and have worked with young children. At each site, the test supervisor will be the same individual who conducts the training of testers as outlined above.

(2) Training of Test Supervisors

The training of test supervisors will consist of a two-day intensive in-service training session conducted at the Juárez and Associates office. The training will orient the test supervisors to the sensitivity needed to test preschool children and to the difficulties involved. It will also acquaint them with the purposes of each test; train them in the training and supervision of testers; and familiarize them with monitoring and problem solving procedures and with the packaging and shipping procedures necessary to ensure the safe arrival of materials at the Juárez and Associates offices. It is expected that most if not all of the supervisors will be selected from current Juárez and Associates staff. These staff members have already been trained in the administration of the test battery and in procedures for the adequate training and supervision of local testers.

(3) Use of Checklists and Systematic Observations of Testers

It is recommended that systematic observations be undertaken by the supervisors of the testers, thereby enabling ratings of the administration of the tests, and relationships of the tester with the children, and the efficient use of time and materials in gathering the necessary data. This procedure will provide feedback,

enabling testers to remain consistent in their performance; identify problems in the administration of the tests, permitting corrective action to be taken; and ensure uniform quality of the data collected.

(4) Daily Review of Testing

It is proposed that at the end of every day of testing a meeting of the supervisor and testers be held to review difficulties or problems which occurred during the day; assess the test booklets for completeness, clarity, and accuracy of notes and responses recorded; and determine that the appropriate coding and scoring procedures have been used. The supervisors and testers will also review each test booklet.

(5) Shipping and Packaging

It is recommended that the test booklets with the response data accompany the supervisor on his or her return to Los Angeles. Other materials, including tape recorders, prop packages, and related matter should be sent by the supervisor via United Parcel Service to Juárez and Associates in Los Angeles.

e. Administration, Coding, and Scoring

(1) It is important that a tape recorder be used in the testing process to verify coding and scoring.

(2) The examiner's manual produced by Juárez and Associates for the pilot test should be revised to reflect the test battery changes and to take into account the background and preparation of the testers.

(3) It is recommended that children not be tested for more than one session per day and that this session should not coincide with snack time or lunch time.

(4) Testing should be conducted in a quiet, comfortable locale where human and material distractions are minimized.

(5) Examiners should be instructed to write legibly, thereby enabling the coding and scoring process to proceed in as reliable and valid a way as possible. This is especially important in determining whether the examiner is reflecting a mispronunciation of a child or an incorrect response or whether the written response is correct but garbled by the poor handwriting of the examiner.

(6) A number of additional recommendations regarding the administration, coding, and scoring of the tests emerged as a result of the pilot testing. These will prove useful in the training and monitoring of the local testers. Examples of these issues include the need for testers to employ culturally and linguistically appropriate and consistent vocabulary at each of the various sites; the need for testers to read test questions carefully and slowly, emphasizing particular words consistently, thereby keeping the children's attention and ensuring reliability; the importance of having practice in both Spanish and English for certain tests and to mount pictures on cards for ease of test administration (i.e., Escuchen Este Cuento); the need to coordinate the relationship between sequencing of answer sheets and test items to allow for the use of examples; the difficulty of coding the responses of children in accord with pre-established categories without distorting the meaning of the responses; the importance of having testers knowledgeable in the grammar of both Spanish and English; the difficulty of coding the responses of children who switch languages and the importance of developing some consistency in this regard; and the need to share beforehand the criteria for scoring all of the tests with the testers so that they can perform the test administration more effectively and efficiently.

Juárez and Associates, Inc.

Appendix A

Letter to Head Start Directors

This letter is to make arrangements, confirm dates, and provide you with information about the procedures to be used in Juárez and Associates' pilot study of Head Start children.

As per our conversation, we will be pilot testing at your Head Start center on _____. Our plan is to be at your site for five consecutive days, during which time we will be supervising the testing of the children. Enclosed is a copy of the timeline we will be following. This timeline describes all of the activities we will be engaged in during the five days we are there.

During our telephone conversation we expressed the need for your cooperation in recruiting local bilingual examiners. We feel it is critical that the people testing the children be familiar with regional language usage. We would appreciate it, therefore, if you would mail to us by _____ a list of six or seven names of individuals who fulfill the requirements of the enclosed job description.

Also, would you please help us in distributing and collecting parent consent forms? We have enclosed thirty copies of these forms. The obtaining of parents' permission to test their children is critical to the pilot testing. We have found through experience that the easiest way to acquire parental permission is to distribute and collect all consent forms during a regularly scheduled parent meeting.

In addition, we will need your help in compiling a list of names of children who will be involved in the pilot testing. At a later data we will be forwarding to you copies of the student data sheet. The student data sheet asks for information on the child's age, sex, date of enrollment, and language preference.

We appreciate your time and effort in helping us do our job better. If you have any questions or comments or encounter any problems with our requests, please do not hesitate to call us at (213) 879-5901. Thank you for your cooperation.

1100 Glendon Avenue • Los Angeles, California 90024 • (213) 478-0826 • (213) 879-5809

Table 2

Testing and Training for Pilot Study

1st Day

a. Interview and screen examiners
b. Complete application forms for employment (W-4)
c. Begin training

2nd day

a. Examiners interact with pilot children
b. Training continues and ends
c. Random assignment of children to examiners
d. Assignment of test sequence to examiners

3rd day

a. Testing of children begins
b. Evaluation team present to monitor performance of examiners
c. At end of testing period examiners
    1. review test booklets for completeness
    2. score PSI, Escuchen Este Cuento, and Listen to the Story

4th day

a. Testing of children continues
b. Evaluation team continues to monitor the examiners
c. At end of testing period, examiners
    1. review test booklets for completeness
    2. score the appropriate tests (see 3rd day)

5th day

a. Testing of children is completed
b. At the end of testing period, examiners
    1. review test booklets for completeness
    2. score appropriate tests (see 3rd day)

c. Evaluation team conducts a debriefing session with all examiners

6th day

a. Pick-up session for children who were absent or have incomplete test batteries.
b. At end of testing period, examiners:
    1. review test booklets for completeness
    2. score appropriate tests (see 3rd day)

7th day

a. Additional pick-up session, if required
b. Examiners rate the success of the testing procedures
c. Examiners mail Juárez and Associates completed tests and test materials

Parental Permission/Permiso de los Padres

I hereby grant permission to Juárez and Associates for my child to take part in the testing program for Head Start children. This testing has been authorized by _____.
name of agency
I understand that my child may either be observed in the classroom or be asked to answer a few questions about the tests. I further understand that the information collected will be used only to find out if these tests are helpful, and that the purpose of Juárez and Associates' work is to improve the education of young children.

Yo le doy permiso a Juárez y Asociados que mi niño/niña participe en un estudio de pruebas para niños de Head Start. Este estudio está autorizado por _____. Yo entiendo que mi
nombre de la agencia
niño/niña quizás sea observado en su salón de clases or posiblemente también se le hagan algunas preguntas sobre las pruebas. La información obtenida solo será usada para el propósito de evaluar la utilidad de las pruebas. Mejorar la educación que reciben los niños es el propósito final de nuestra labor.

_____ (child's name) (nombre del nino)

_____ (signature of parent or guardian) (firma del padre o guardián)

_____ (date) (fecha)

Your permission and cooperation are appreciated.

Su permiso y cooperación le serán muy agradecidas.

Code # _____

STUDENT DATA SHEET

We are asking each teacher involved with the Head Start Bilingual/Bicultural Curriculum Development Project to complete this data sheet for each child's background. It should be returned to Juárez and Associates by _____. Each question is to be answered as accurately as possible. The data sheet should be filled out in time to be included with completed tests. Most of the information asked for in this data sheet has already been collected by Head Start personnel.

CHILD'S BACKGROUND: TO BE COMPLETED BY TEACHER

1. Name of child: _____

   Sex: M _____ F _____ Birthdate _____

   Age: Years_____ Months_____

   Where was child born: City_____ State_____

2. Principal language spoken before enrollment in Head Start?

   English_____ Spanish_____ Both_____

   Which language does this child use most often with best friends?

   English_____ Spanish_____ Both_____

   Which language does this child use most often on playground?

   English_____ Spanish_____ Both_____

   Which language does this child use most often with teacher?

   English_____ Spanish_____ Both_____

   Which language is most often spoken at home?

   English_____ Spanish_____ Both_____

   Language Preference: English_____ Spanish_____ Both_____

3. How long has child been enrolled in this Head Start program?

   _____

4. Have any of this child's older siblings been enrolled in this Head Start?

   Yes_____ No_____ If yes, how many? _____

5. How much time does the child spend in the Head Start Center each day?

   Indicate hours _____

6. How many hours a week does the child spend in the center?

   Indicate hours_____

7. Is child enrolled in the morning or afternoon session (please circle)?

   Morning Afternoon

8. Total number of children enrolled at this Head Start Center?

   _____

9. The first session runs from _____ a.m. to _____ p.m.

   The second session runs from _____ a.m. to _____ p.m.

CORPUS CHRISTI (01)

CODES	AGE	SEX	ETHNIC BACKGROUND	LANGUAGE PREFERENCE
01-11-001	5 years (60 months)	F	Mexican American	English
01-11-002	4 years 9 months (57 months)	F	Mexican American	Spanish
01-11-003	4 years 8 months (56 months)	M	Mexican American	English
01-11-004	5 years 3 months (63 months)	M	Mexican American	English
01-11-005	4 years 7 months (55 months)	M	Mexican American	English
01-12-006	5 years 1 month (61 months)	F	Mexican American	English
01-12-007	5 years 1 month (61 months)	F	Mexican American	English
01-12-008	4 years 4 months (52 months)	M	Mexican American	English
01-12-009	5 years 3 months (63 months)	F	Mexican American	English
01-12-010	5 years 2 months (62 months)	F	Mexican American	English
01-13-011	4 years 11 months (59 months)	F	Mexican American	English
01-13-012	4 years 4 months (52 months)	M	Mexican American	English
01-13-014	4 years 4 months (52 months)	F	Mexican American	English
01-13-015	4 years 8 months (56 months)	F	Mexican American	English
01-14-016	5 years 5 months (65 months)	F	Mexican American	English
01-14-017	5 years 2 months (62 months)	M	Mexican American	Spanish
01-14-018	4 years 9 months (57 months)	M	Mexican American	Spanish
01-14-019	5 years 2 months (62 months)	F	Mexican American	English
01-14-020	4 years 10 months (58 months)	F	Mexican American	English
01-15-021	5 years 5 months (65 months)	F	Mexican American	English
01-15-022	5 years 3 months (63 months)	M	Mexican American	English
01-15-023	5 years 5 months (65 months)	F	Mexican American	English
01-15-024	4 years 10 months (58 months)	F	Mexican American	English
01-15-025	4 years 8 months (56 months)	F	Mexican American	English

RIO GRANDE (02)

CODES	AGE	SEX	ETHNIC BACKGROUND	LANGUAGE PREFERENCE
02-21-026	4 years 10 months (58 months)	F	Mexican American	Spanish
02-21-027	5 years 1 month (61 months)	M	Mexican American	Spanish
02-21-028	5 years 1 month (61 months)	F	Mexican American	Spanish
02-21-029	4 years 9 months (57 months)	F	Mexican American	Spanish
02-22-030	4 years 8 months (56 months)	M	Mexican American	Spanish
02-22-031	5 years 5 months (65 months)	M	Mexican American	Spanish
02-22-032	4 years 7 months (55 months)	F	Mexican American	Spanish
02-22-033	5 years 6 months (66 months)	M	Mexican American	Spanish
02-22-034	4 years 7 months (55 months)	M	Mexican American	Spanish
02-23-035	5 years 2 months (62 months)	F	Mexican American	Spanish
02-23-036	4 years 6 months (54 months)	M	Mexican American	Spanish
02-23-037	4 years 6 months (54 months)	M	Mexican American	Spanish
02-23-038	4 years 9 months (57 months)	M	Mexican American	Spanish
02-23-039	4 years 9 months (57 months)	M	Mexican American	Spanish
02-24-040	5 years 3 months (63 months)	F	Mexican American	Spanish
02-24-041	4 years 9 months (57 months)	F	Mexican American	Spanish
02-24-042	4 years 6 months (54 months)	F	Mexican American	Spanish
02-24-043	4 years 9 months (57 months)	M	Mexican American	Spanish
02-24-044	4 years 9 months (57 months)	M	Mexican American	English
02-25-045	5 years 5 months (65 months)	F	Mexican American	Spanish
02-25-046	4 years 6 months (54 months)	F	Mexican American	Spanish
02-25-047	5 years 4 months (64 months)	M	Mexican American	Spanish
02-25-048	5 years 1 month (61 months)	M	Mexican American	Spanish
02-25-049	5 years 2 months (62 months)	F	Mexican American	Spanish
02-21-050	4 years 4 months (52 months)	F	Mexican American	Spanish

02-21-050	4 years 4 months (52 months)	F	Mexican American	Spanish

MILWAUKEE (03)

CODES	AGE	SEX	ETHNIC BACKGROUND	LANGUAGE PREFERENCE
03-31-052	4 years 2 months (50 months)	M	Mexican American	English
03-31-053	4 years (48 months)	F	Anglo	English
03-31-054	4 years 4 months (52 months)	F	Mexican American	Spanish
03-31-055	4 years 8 months (56 months)	F	Anglo	English
03-31-056	4 years 8 months (56 months)	F	Puerto Rican	English
03-31-057	4 years 6 months (54 months)	M	Mexican American	Spanish
03-32-058	5 years 1 month (61 months)	M	Mexican American	Spanish
03-32-059	3 years 7 months (43 months)	M	Mexican American	Spanish
03-32-060	4 years 9 months (57 months)	M	Mexican American	Spanish
03-32-061	4 years 4 months (52 months)	M	Anglo	English
03-33-062	5 years 2 months (62 months)	F	Mexican American	English
03-33-063	5 years 2 months (62 months)	F	Mexican American	English
03-33-064	5 years 2 months (62 months)	M	Mexican American	English
03-33-065	4 years 6 months (54 months)	M	Puerto Rican	Spanish
03-33-066	4 years 7 months (55 months)	F	Mexican American	English
03-34-067	4 years 2 months (50 months)	F	Mexican American	Spanish
03-34-068	4 years 5 months (53 months)	F	Mexican American	English
03-35-069	3 years 11 months (47 months)	F	Mexican American	English
03-34-070	4 years 3 months (51 months)	F	Mexican American	English
03-34-071	5 years (60 months)	F	Anglo	English

NEW YORK (04)

CODES	AGE	SEX	ETHNIC BACKGROUND	LANGUAGE PREFERENCE
04-41-073	4 years 11 months (59 months)	F	Black	English
04-41-074	4 years 6 months (54 months)	M	Puerto Rican	English
04-42-075	4 years 3 months (51 months)	F	Dominican	Spanish
04-42-077	5 years (60 months)	M	Dominican	English
04-42-078	4 years 10 months (58 months)	M	Puerto Rican	English
04-43-081	4 years 8 months (56 months)	F	Puerto Rican	Both
04-43-082	5 years (60 months)	M	Puerto Rican	English
04-43-083	4 years 10 months (58 months)	M	Dominican	Spanish
04-43-084	4 years 9 months (57 months)	M	Black	English
04-43-085	4 years 6 months (54 months)	F	Black	English
04-44-087	4 years 5 months (53 months)	M	Chinese	English
04-44-088	5 years 2 months (62 months)	F	Puerto Rican	English
04-44-089	5 years 11 months (71 months)	M	Puerto Rican	Both
04-44-090	4 years 6 months (54 months)	F	Black	English
04-45-092	5 years (60 months)	M	Puerto Rican	English
04-45-093	4 years 10 months (58 months)	M	Puerto Rican	English
04-45-094	4 years 4 months (52 months)	F	Puerto Rican	English
04-45-095	5 years 3 months (63 months)	F	Black	English
04-46-098	5 years (60 months)	F	Puerto Rican	Both
04-46-099	5 years 2 months (62 months)	M	Puerto Rican	Spanish
04-46-100	4 years 4 months (52 months)	F	Puerto Rican	English

Job Description

Test Examiners for the Pilot Study of the Assessment Instruments

Local examiners will be needed at each pilot site to administer the evaluation test battery. These examiners will be responsible for administering the entire test battery in both Spanish and English. Because many of the test measures assess various aspects of receptive and expressive language, it is necessary that the examiner be bilingual and biliterate. The potential examiner should meet the following criteria:

I. Educational Background

A minimum of a high school degree is preferred. Formal training in child development or related area is desirable.

II. Experience

The individual should have experience in working with young Spanish-speaking children.

III. Responsibilities

1. Administer entire test battery.
2. Collect student profile data: i.e., age, sex, language preference.

IV. Requirements

1. The individual should be able to converse fluently in Spanish and English.
2. The individual must be able to read in Spanish and English.
3. The individual must be able to write down a story as told by a child in Spanish and English. Further, listening comprehension of what the child is saying in both languages is important.
4. The "speech pattern" in Spanish (pronunciation, language rhythm, etc.) of the individual should be compatible with the "speech pattern" of the children being tested.

V. Salary

$3.50 to $4.50 per hour.

VI. Approximately 10 working days

Pilot study examiners have the option of working in the Fall of 1979 and the Spring of 1980 for a longer period of time.

Tester's Name_____
(code)

Observer's Name_____ Date_____

Site:_____

Name of Test_____

		Yes	No
I.	Instructions		
	a. Are the questions read clearly and audibly?	( )	( )
	b. Does the examiner emphasize key words according to the directions in the Tester's Manual (e.g., as for the productive language measures)?	( )	( )
	c. Does the examiner keep his/her voice at an even volume level when reading questions to the child? (Changes in the inflection of the examiner's voice may serve as cues for the child.)	( )	( )
II.	Reinformcement		
	a. Does the examiner verbally reward the child for his/her response regardless of the correctness of the response?	( )	( )
	b. Does the examiner differentially reward the child for correct answers?	( )	( )
	c. Does the examiner unknowingly cue the child to the correct answer?	( )	( )
III.	Rapport		
	a. Does the examiner act in a warm and friendly manner toward the child?	( )	( )
	b. Is the examiner patient with the child during the testing session?	( )	( )
	c. Does the examiner appear to be patient with the child throughout the testing session?	( )	( )

Appendix E con't

	Yes	No
d. Is the examiner able to keep the child on the task?	( )	( )
e. Is the examiner able to keep the child moving along on the task?	( )	( )

IV. Time

a. Does the examiner allow the child to waste time (by attending to irrelevant questions and requests)?	( )	( )
b. Does the examiner appear to rush the child through the test?	( )	( )
c. Does the examiner pace his/her questions evenly and at a constant rate?	( )	( )

Comments

VI. Bibliography

Burt, M.K., Dulay, H.C., and Hernandez-Ch., E. Bilingual syntax measure technical handbook. New York: Harcourt Brace Jovanovich, Inc., 1975.

Cooperative Preschool Inventory: Spanish Edition. Princeton, New Jersey: Educational Testing Service, 1974.

Deloria, D., et al. The National Home Start Evaluation Interim Report V. Summative Evaluation Results. Ypsilanti: High/Scope Educational Research Foundation, 1974.

Hardy, R., Glad, D., Bernal, E. M., and Cordova, F. Circo technical report and manual, Princeton, N.J.: Educational Testing Service, 1978.

Juárez and Associates. Phase II: A plan for the pilot study of child and parent impact measures. Los Angeles: December 1978.

Juárez and Associates. Review and recommendation for the test battery. Los Angeles: July 1978.

Preschool Inventory: Revised Edition. New York: Addison-Wesley Publishing Company, 1970.

Stallings, J., Wilcox, M., and Travers, J. Final report on phase II instruments for the national day-care cost effects study: Instrument selection and field testing. Menlo Park, California: Stanford Research Institute, 1976.